

Yan *et al*

1

2 A comprehensive foundation model for cryo-EM image

3 processing

4

Yang Yan^{1,2,6}, Shiqi Fan^{1,2,6}, Fajie Yuan^{1,2,*}, and Huaizong Shen^{1,3,4,5,*}

6

7

⁸ ¹Research Center for Industries of The Future (RCIF), Westlake University, Hangzhou, Zhejiang,
⁹ China

10

11

¹²School of Engineering, Westlake University, Hangzhou, Zhejiang, China

13

14

¹⁵ ³Zhejiang Key Laboratory of Structural Biology, School of Life Sciences, Westlake University;
¹⁶ Hangzhou, Zhejiang, China

17

18

19 ⁴Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang, China

20

21

22 ⁵Institute of Biology, Westlake Institute for Advanced Study, Hangzhou, Zhejiang, China

23

24

²⁵ ⁶These authors contributed equally to this work.

26

27

²⁸*Correspondence: F. Yuan (yuanfajie@westlake.edu.cn), H. Shen
²⁹(shenhuazong@westlake.edu.cn)

31 **Abstract**

32 **Cryogenic electron microscopy (cryo-EM) has become a premier technique for high-**
33 **resolution structural determination of biological macromolecules¹⁻⁴. However, its**
34 **widespread adoption is hampered by the need for specialized expertise. We introduce**
35 **the Cryo-EM Image Evaluation Foundation (Cryo-IEF) model, pre-trained on an**
36 **extensive dataset of approximately 65 million cryo-EM particle images using**
37 **unsupervised learning. Cryo-IEF excels in various cryo-EM data processing tasks,**
38 **such as classifying particles from different structures, clustering particles by pose, and**
39 **assessing the quality of particle images. Upon fine-tuning, the model effectively ranks**
40 **particle images by quality at high efficiency, enabling the creation of CryoWizard—a**
41 **fully automated single-particle cryo-EM data processing pipeline. This pipeline has**
42 **successfully resolved high-resolution structures of diverse properties and proven**
43 **adept at mitigating the prevalent preferred orientation challenge in many cryo-EM**
44 **samples. The Cryo-IEF model and CryoWizard pipeline collectively represent a**
45 **significant advancement in rendering cryo-EM technology more accessible, efficient,**
46 **and robust, with substantial implications for life sciences research.**

47

48 **Main**

49 Cryogenic electron microscopy (cryo-EM) has become an indispensable tool for the
50 structural analysis of biological macromolecules¹⁻⁴. Unlike crystallography, cryo-EM does
51 not require crystallization, which is often a significant obstacle for many biological
52 specimens⁵. The past decade has witnessed a transformative leap in cryo-EM⁶⁻⁸, particularly
53 in single-particle analysis, driven by advancements in direct detector devices and
54 computational methods⁹⁻¹⁴, establishing it as a leading technique for structural elucidation.

55

56 Despite its potential, the broader application of cryo-EM is limited by the intricate
57 data processing required, posing a substantial barrier to non-experts^{15,16}. A typical cryo-EM
58 workflow involves multiple steps, each traditionally requiring manual oversight, which can
59 be labor-intensive, time-consuming, and prone to human error, potentially leading to
60 suboptimal outcomes. To address these challenges, deep learning-based methods have been
61 developed to automate specific workflow segments¹⁷⁻²³. However, these solutions tend to
62 be task-specific and still require considerable user intervention. Moreover, challenges such
63 as structural heterogeneity^{21,24} and preferred orientation²⁵⁻²⁷ continue to complicate
64 structure determination from many biological samples, despite numerous attempted
65 remedies.

66

67 Pre-trained vision foundation models have proven extremely effective for a
68 variety of image processing tasks, often outperforming traditional training methods on
69 specialized datasets²⁸⁻³¹. These models represent a new standard in image processing, with
70 broad applications across various domains, including medical image analysis³²⁻³⁵ and the

71 processing of fluorescence microscopy images³⁶. Within the realm of cryo-EM data
72 processing, several pivotal steps, such as particle pose estimation and 2D/3D classification,
73 could potentially benefit from a pre-trained foundation model tailored to particle images.
74 However, to the best of our knowledge, no such foundation models have yet been
75 introduced.

76

77 In this study, we introduce the cryo-EM Image Evaluation Foundation Model
78 (Cryo-IEF), tailored for a broad range of cryo-EM image processing tasks. The model
79 demonstrates superior performance in the classification of particles from different
80 structures, the clustering of particles by pose, and the assessment of particle quality.
81 Building on this foundation model, we have developed CryoWizard, a fully automated
82 pipeline for cryo-EM data processing, which further streamlines the cryo-EM workflow.

83

84 **An overview of Cryo-IEF**

85 Cryo-IEF employs a contrastive learning framework akin to MoCo v3 to learn feature
86 representations from cryo-EM particle images (Fig. 1)³⁰. Each particle image in the pre-
87 training dataset is processed to create two separate, independently augmented views, which
88 are then encoded using two parallel encoders. A variety of data augmentation techniques
89 are employed, such as random cropping, color jittering, blurring, solarization, and rotation.
90 The core objective during training is to enhance the similarity of feature representations
91 between different views of the same particle image while reducing similarity between
92 views of distinct particle images. For detailed model architecture descriptions and training
93 processes, please refer to the Methods (Extended Data Fig. 6a,b).

94

95 The foundation model was trained using an extensive dataset comprising
96 approximately 65 million cryo-EM particle images, which included over a hundred distinct
97 biological structures. This diverse collection of datasets was amassed from several sources:
98 the Electron Microscopy Public Image Archive (EMPIAR) database³⁷, a previous study³⁸,
99 and our own in-house experimental data, with comprehensive details provided in the
100 Methods section (Extended Data Fig. 7 and Extended Data Tables 1 and 2).

101

102 To rigorously assess the performance of the Cryo-IEF model across a broad
103 spectrum of cryo-EM image processing tasks, we conducted semantic representation
104 analysis on the Cryo-IEF-extracted features from test datasets not used in model training.
105 These datasets include 12 simulated and 4 genuine particle datasets (Tables 1 and 2). The
106 simulated datasets were generated from high-resolution cryo-EM maps downloaded from
107 the Electron Microscopy Data Bank (EMDB) and were labeled with ground-truth
108 information regarding structural variations and pose positions (Table 1). The molecular
109 weights of the related structures range from 69 kDa to 1 MDa for the simulated datasets and
110 from 68 kDa to 2.2 MDa for the genuine datasets. The details of these test datasets can be
111 found in the Methods.

112

113 **Cryo-IEF classifies particles from different structures**

114 Efficient and reliable classification of particles from different structures is crucial in cryo-
115 EM data processing²⁴, as it not only allows the resolving of each structure, but also

116 improves the throughput of structural determination, potentially enabling resolving multiple
117 structures in one collected dataset.

118

119 Recent studies have demonstrated that many pre-trained vision foundation models
120 possess three dimensional awareness³⁹. To test whether Cryo-IEF also has such capability
121 and classifies particles from different structures, we conducted an initial assessment using
122 k-Nearest Neighbors (k-NN) classification on test dataset particles. The features extracted
123 by Cryo-IEF from 12 simulated datasets and 4 genuine datasets exhibited distinct clustering
124 behaviors based on structural variations, achieving k-NN accuracies of 96.42% and
125 93.32%, respectively (Fig. 2). These results demonstrate that Cryo-IEF effectively
126 differentiates particles from distinct structures.

127

128 To further explore Cryo-IEF's capability in classifying particles from different
129 structures, we applied it to deep learning-based heterogeneous structural reconstruction
130 tasks. Existing methods for such tasks often employ an autoencoder-like framework, where
131 the structural reconstruction involves two main steps: an encoder projects particle images
132 into a feature space, and a decoder reconstructs the 3D structure from this space^{21,40-42}.
133 Typically, both the encoder and decoder are randomly initialized and trained together
134 iteratively—the encoder to capture structural heterogeneity and the decoder to represent
135 these structures in three dimensions.

136

137 In our study, we introduced a novel heterogeneous structural reconstruction AI
138 model named CryoSolver (Fig. 3a). This model integrates Cryo-IEF as a fixed feature

139 extractor within the DRGN-AI framework, replacing the standard trainable encoder⁴².
140 Unlike traditional approaches, Cryo-IEF remains frozen with fixed parameters, focusing
141 solely on feature extraction, while the decoder remains trainable. We evaluated CryoSolver
142 using a resampled version of the Ribosemby dataset from CryoBench, which includes four
143 distinct structures (Fig. 3b)⁴³. The dataset features particles with varying proportions:
144 12.75% for structure 8C9C, 62.32% for 8C99, 16.82% for 8C93, and 8.11% for 8C8X.
145 Detailed dataset information and training processes are provided in the Methods section. In
146 contrast to the distribution of randomly initialized features without any clustering (Fig.
147 3c), the Cryo-IEF-extracted features from the test dataset particles were well-clustered, with
148 particles from the same structure closely distributed (Fig. 3d). Following the hierarchical
149 pose search process and stochastic gradient descent training of the decoder, we randomly
150 selected three points for each structure from the feature space to perform 3D reconstruction.
151 The reconstructions based on randomly initialized features predominantly resembled the
152 second ground-truth structure (PDB ID: 8C99), which was the majority in the dataset
153 (62.32%) (Fig. 3e). On the contrary, the structures generated by CryoSolver closely
154 matched the ground-truth 3D structures (Fig. 3f). These findings indicate that Cryo-IEF,
155 despite being trained in an unsupervised manner without explicit structural heterogeneity
156 labels, effectively extracts features that reflect 3D structural variations, thereby
157 demonstrating its capability to classify particles from different structures.

158

159 **Cryo-IEF clusters particles by pose**

160 In cryo-EM, the pose of a particle refers to its orientation and position within the context of
161 its three-dimensional reconstruction. Accurate pose determination is crucial for achieving

162 high-resolution cryo-EM maps^{13,14,44}. Additionally, many cryo-EM samples suffer from the
163 preferred orientation problem^{25-27,45}, where most particles of a dataset adopt a limited
164 number of dominant poses, leading to incomplete structural information. Accurately
165 determining the pose of each particle can help reduce these preferred orientations, leading
166 to more accurate reconstructions.

167

168 The Cryo-IEF model is trained to maximize its ability to identify unique particle
169 features, which suggests that Cryo-IEF can cluster particles with similar poses. To test this
170 hypothesis, we evaluated the model's performance in pose clustering using 12 simulated
171 datasets, by employing the k-NN classification. Indeed, Cryo-IEF-extracted particle
172 features demonstrated distinct clustering patterns that matched the ground-truth pose
173 classes, achieving a mean k-NN accuracy of 90.6% (Fig. 4). The model showed exceptional
174 performance for large molecular weight samples (over 400 kDa), with k-NN scores
175 exceeding 95%. These results confirm that Cryo-IEF can effectively cluster cryo-EM
176 particles by their poses. Further processing was conducted for the dataset EMD-24057. The
177 particles of each cluster in EMD-24057 underwent 2D classification in CryoSPARC¹⁴,
178 which revealed clear 2D class average images. Remarkably, the adjacent clusters visualized
179 in the Uniform Manifold Approximation and Projection (UMAP)⁴⁶ plot corresponded to
180 particles with similar 2D class average images, indicating that the spatial distribution of
181 features extracted by Cryo-IEF reflects the relationship between different particle poses,
182 further demonstrating its effectiveness in pose clustering.

183

184 **Cryo-IEF distinguishes particle quality**

185 The primary objective of cryo-EM data processing is to achieve high-resolution structures
186 by accurately aligning particle images⁴⁷. This requires selectively retaining high-quality
187 particles while discarding those that could compromise the resolution^{15,44,48,49}. Here,
188 particle quality is defined by the extent to which a particle contributes to the overall quality
189 and resolution of the final constructed cryo-EM maps. Therefore, distinguishing and
190 ranking particle images by quality is crucial for successful reconstruction. Cryo-EM images
191 typically have extremely low signal-to-noise ratios (SNR), making it challenging to discern
192 the image quality^{50,51}. Consequently, Current algorithms cannot evaluate images
193 individually; instead, particles with similar views are clustered through 2D classification to
194 create averaged images, which are then collectively retained or discarded^{13,14}. This
195 approach, however, requires manual intervention and can introduce errors in decision-
196 making on single images.

197

198 To tackle this challenge, we evaluated Cryo-IEF's ability to assess cryo-EM image
199 quality on a per-particle basis. We applied the k-NN algorithm to four genuine test datasets.
200 Initially, particles underwent 2D classification in CryoSPARC, and each class was
201 manually assigned a quality score based on the clarity of the class averages, ranging from 0
202 (pure noise or ice) to 1 (clear structural features) (Extended Data Fig. 1). Based on these
203 scores, particles were categorized into high-quality (above 0.7), medium-quality (between
204 0.3 and 0.7), and low-quality (below 0.3) classes. Cryo-IEF-extracted features displayed
205 distinct clustering with a k-NN accuracy of 74.06% (Fig. 5b), indicating that Cryo-IEF can
206 differentiate particle quality without explicit training.

207

208 To improve Cryo-IEF's capability in distinguishing and ranking particle quality,
209 we developed CryoRanker, a fine-tuned version of Cryo-IEF, which incorporates Cryo-
210 IEF's backbone encoder with an additional classifier head (Fig. 5a and Extended Data Fig.
211 6c). A combination of cross-entropy and margin ranking losses were employed for fine-
212 tuning. CryoRanker was trained on a comprehensive dataset of approximately 42 million
213 labeled cryo-EM particle images, the quality scores of which were assigned using the same
214 criteria as for the test datasets. Details of the training datasets and process are provided in
215 the Methods section (Extended Data Fig. 7). Features extracted by CryoRanker from the
216 four genuine test datasets showed clearer clustering with a k-NN score of 76.65%,
217 surpassing Cryo-IEF's performance (Fig. 5c). The predicted scores by CryoRanker closely
218 matched the average labeled scores (Fig. 5d), supporting the ranking ability of the model.
219 CryoRanker was also evaluated with precision and recall metrics, which reveal higher
220 values for high- and low-quality particles than medium-quality ones (Extended Data Fig.
221 2). It is worth noting that CryoRanker exhibits high processing efficiency. In our tests,
222 ranking one million particles required only an average of 35 minutes when using four
223 NVIDIA V100 GPUs.

224

225 **An automated single-particle cryo-EM data processing pipeline**

226 Multiple rounds of 2D and 3D classifications in cryo-EM data processing are resource-
227 intensive, time-consuming, and often require human intervention. To streamline these
228 processes, we developed CryoWizard, a fully automated cryo-EM data processing pipeline,
229 by integrating CryoRanker to replace the 2D/3D classification steps (Fig. 6a and Extended

230 Data Fig. 3). Currently, CryoWizard is designed only for homogeneous structure
231 reconstruction.

232

233 CryoWizard comprises three main stages which are data preprocessing and particle
234 quality ranking, initial model reconstruction, and final map refinement (Fig. 6a and
235 Extended Data Fig. 3). First, collected cryo-EM movies or micrographs are preprocessed to
236 produce extracted particles, which are ranked and assigned quality scores by CryoRanker.
237 Second, optimal number of top-ranked particles (e.g., 50,000) are selected to generate an
238 initial model, based on which a refined map is reconstructed as the template for
239 downstream structural refinement. Third, a larger optimal number of top-ranked particles
240 (e.g. 200,000) are chosen for the final high-resolution map refinement. For datasets
241 afflicted by preferred orientation issues, an additional clustering module is implemented in
242 the pipeline to address this problem, as detailed in the next section. Except for the
243 assignment of quality scores by CryoRanker, the remaining tasks in the current version of
244 CryoWizard are conducted in CryoSPARC by interfacing with cryosparc-tools¹⁴.

245

246 The performance of CryoWizard were evaluated with six datasets downloaded
247 from the Electron Microscopy Public Image Archive (EMPIAR): EMPIAR-10556,
248 EMPIAR-11292, EMPIAR-10405, EMPIAR-10217, EMPIAR-10250, and EMPIAR-10876
249 (Fig. 6b-g). These datasets were not used in model training. The final refined maps,
250 automatically resolved by the pipeline without performing any postprocessing, local or CTF
251 refinement, or Bayesian polishing, achieved resolutions of 2.88 Å, 2.62 Å, 3.24 Å, 2.39 Å,
252 3.58 Å, and 3.01 Å, respectively. These resolutions are sufficient for accurate model

253 building. The corresponding models of these datasets were downloaded, which align well
254 with the resolved density maps, confirming their correctness.

255

256 **CryoWizard addresses the preferred orientation problem**

257 The preferred orientation problem in many cryo-EM samples arises when most particles
258 adopt a limited number of dominant poses, often due to surface charges or other unknown
259 factors^{25-27,45}. We have previously shown that Cryo-IEF can effectively cluster particles by
260 their poses. This led us to hypothesize that certain particle classes identified by Cryo-IEF
261 may contain particles less affected by preferred orientation, which could be used to
262 generate isotropic maps with uniform resolutions in all directions. Therefore, we
263 implemented an additional clustering module within CryoWizard which is triggered by the
264 presence of preferred orientation problem (Fig. 6a and Extended Data Fig. 3).

265

266 We tested CryoWizard's capability to address the preferred orientation problem on
267 two datasets: EMPIAR-10217 and EMPIAR-10096, which exhibited highly imbalanced
268 pose distributions with only a few dominant poses, as shown in the 2D classification results
269 (Extended Data Figs. 4 and 5). Conical FSC Area Ratio (cFAR) score, which ranges from 0
270 to 1, evaluates the severity of the preferred orientation problem in cryo-EM maps—a lower
271 score indicates a more severe issue⁴⁵. Routine processing of EMPIAR-10217 and EMPIAR-
272 10096 datasets with manually selected particles resulted in maps severely impacted by
273 preferred orientation, with cFAR scores of 0.01 and 0.03, respectively (Fig. 7b,d and
274 Extended Data Figs. 4 and 5). However, by clustering the Cryo-IEF-extracted particle
275 features of EMPIAR-10217 using K-Means++ algorithm, we identified a particle class that

276 produced a map with similar resolutions in all directions, achieving a cFAR score of 0.75
277 (Fig. 7a). Using this map as a template, the final refined map was exempt from preferred
278 orientation effects and achieved a resolution of 2.37 Å (Fig. 7c). Similarly, the CryoWizard
279 pipeline effectively addressed the preferred orientation issue in EMPIAR-10096 dataset
280 (Fig. 7e).

281

282 **Discussion**

283 In this study, we introduced Cryo-IEF, the first pre-trained foundation model specifically
284 designed for cryo-EM particle image processing. Utilizing a large-scale, unsupervised
285 model trained on approximately 65 million images, Cryo-IEF demonstrates the
286 transformative potential of deep learning in cryo-EM data processing. Despite its
287 unsupervised pre-training, Cryo-IEF effectively distinguishes particles based on structural
288 variations, pose angles, and particle quality, achieving high k-NN accuracies and
289 demonstrating three-dimensional awareness.

290

291 The development of CryoRanker, a fine-tuned variant of Cryo-IEF, further
292 enhances the model's performance in assessing particle quality. CryoRanker ranks particles
293 individually, offering a significant improvement over traditional methods. Building on
294 these capabilities, we developed CryoWizard, an automated pipeline that integrates
295 CryoRanker to eliminate the need for manual oversight in cryo-EM data processing.
296 CryoWizard effectively automates homogeneous structural reconstruction and demonstrates
297 broad applicability by achieving high-resolution reconstructions across multiple datasets.
298 By reducing consumed computational resources and manual intervention, CryoWizard

299 improves both the efficiency and robustness of cryo-EM structural determination,
300 significantly streamlining the processing pipeline.

301

302 CryoWizard also addresses the preferred orientation problem by identifying
303 particle classes less affected by dominant poses. The pipeline significantly improves map
304 isotropy, as demonstrated in datasets like EMPIAR-10217 and EMPIAR-10096. By
305 achieving isotropic resolutions, CryoWizard provides a robust solution to a longstanding
306 challenge in cryo-EM data processing.

307

308 Despite these advancements, our study has some limitations. First, although Cryo-
309 IEF was trained on an extensive dataset encompassing over a hundred different biological
310 macromolecules, samples of poor quality or low molecular weights (< 100 kDa) are
311 underrepresented. Consequently, the model's performance deteriorates on such samples.

312 Expanding the training dataset to include more of these underrepresented samples will help
313 alleviate this issue. Second, the current version of CryoWizard is designed solely for
314 homogeneous structural reconstruction. We plan to incorporate Cryo-IEF's classification
315 capabilities to handle multiple structures or conformations, thus broadening its applicability
316 and increasing the pipeline's throughput. Third, we aim to optimize CryoWizard further to
317 enhance its efficiency and robustness, and to explore the integration of additional tools to
318 streamline the cryo-EM workflow.

319

320 Overall, Cryo-IEF and its applications exemplify the potential of integrating
321 advanced AI models into cryo-EM workflows. They represent a substantial advancement in

322 rendering cryo-EM technology more accessible, efficient and robust. By reducing
323 consumed computational resources and manual intervention, CryoWizard has the potential
324 to accelerate discoveries in structural biology and extend the applicability of cryo-EM
325 across diverse research domains in biology and beyond.

326

327 **Methods**

328 **Datasets for pre-training and fine-tuning**

329 The datasets for pre-training and fine-tuning were gathered from various sources, including
330 the EMPIAR database³⁷, cryoPPP³⁸, and in-house experiments (Extended Data Fig. 7). For
331 most datasets, we downloaded the movies or micrographs and imported them into
332 CryoSPARC¹⁴. In CryoSPARC, we performed motion correction (Patch Motion
333 Correction; this step is skipped for micrograph data), CTF estimation (CTFFIND4⁵²),
334 particle picking (Blob Picker), and particle extraction using the tools provided by
335 CryoSPARC (version 4.6.0). The extracted particle images were subsequently added to the
336 pre-training dataset. For the fine-tuning dataset, we conducted particle clustering using the
337 2D Classification function in CryoSPARC. Based on the clarity of the 2D class averages,
338 the cryo-EM particle images were manually assigned quality scores, ranging from 0 (pure
339 noise or ice) to 1 (clear structural features) (Extended Data Fig. 1). Particles in the same
340 class were assigned the same quality score. For datasets containing extracted particle
341 images, data preprocessing step was skipped. These images were directly added to the pre-
342 training dataset and imported into CryoSPARC for 2D classification and manually assigned
343 quality scores before being added to the fine-tuning dataset.
344 In total, we gathered 65,310,474 cryo-EM particle images for pre-training the cryo-IEF
345 model and 42,388,319 quality-labeled cryo-EM particle images for fine-tuning
346 CryoRanker. The fine-tuning dataset is a subset of the pre-training dataset due to the
347 exclusion of datasets with poor clustering quality (Extended Data Fig. 8) and the effort to
348 balance the number of high- and low-quality particles.

349

350 **EMPIAR datasets**

351 The EMPIAR IDs used for pre-training and fine-tuning are summarized (Extended Data
352 Table 1). For each EMPIAR ID marked with an asterisk, we were able to download the
353 particle images directly and imported them into CryoSPARC for 2D classification and
354 assignment of quality scores. The downloaded particle images were incorporated into the
355 pre-training dataset, while the manually-labeled particle images were included in the fine-
356 tuning dataset. For the other EMPIAR IDs, we downloaded the movies or micrographs and
357 imported them into CryoSPARC, where we processed the data as described above. The
358 extracted particle images were subsequently added to the pre-training dataset, while the
359 particle images with manually labeled quality scores were included in the fine-tuning
360 dataset.

361

362 **CryoPPP datasets**

363 CryoPPP comprises 34 representative protein datasets selected from EMPIAR. We
364 downloaded all 34 datasets provided by the CryoPPP database. The downloaded particle
365 images were imported into CryoSPARC, where we performed 2D classification. Five
366 datasets could not be imported successfully and were therefore excluded. All 29
367 successfully imported datasets are listed in Extended Data Table 2. The quality of the
368 particle images was evaluated manually based on the 2D averages. All 29 successfully
369 imported particle image datasets were added to our pre-training dataset for cryo-IEF, while
370 25 datasets with manually labeled quality scores were included for fine-tuning of
371 CryoRanker. Five datasets, including EMPIAR-10061, EMPIAR-10345, EMPIAR-10387,
372 EMPIAR-10590, and EMPIAR-10947, were excluded from fine-tuning due to poor

373 averages quality. Based on the 2D classification averages (Extended Data Fig. 8), the good
374 particles could not be distinguished manually, preventing them from providing reliable
375 scores.

376

377 **In-house datasets**

378 We gathered 50 cryo-EM image datasets from in-house experiments that are not available
379 in public databases such as EMPIAR. Similar to the EMPIAR datasets, these datasets
380 consist of movies or micrographs and were processed using the tools provided by
381 CryoSPARC. The processing pipeline mirrors that of the EMPIAR datasets. The extracted
382 particle images were added to the pre-training dataset for Cryo-IEF training, while the
383 particle images with manually labeled quality scores were included in the fine-tuning
384 dataset for CryoRanker training.

385

386 **Datasets for evaluation**

387 **Datasets for Cryo-IEF pre-training evaluation**

388 We downloaded 12 distinct biological particle density maps reconstructed using single-
389 particle cryo-electron microscopy (cryo-EM) from the EMDB database⁵³. These density
390 maps were classified into four categories based on their overall molecular weights (Table
391 1), which were used to generate simulated cryo-EM particle images for evaluation. For each
392 density map, we generated 2,000,000 simulated cryo-EM particle images using the
393 Simulate Data function in CryoSPARC. The signal-to-noise ratio was set to 0.005, while
394 other settings remained at their defaults. Based on the ground-truth pose angles recorded in
395 the metadata generated by CryoSPARC, the simulated cryo-EM particle images were

396 classified into 12 classes. We sampled 2,000 images for each pose angle class. To avoid
397 ambiguity in class division, only the 50% of particles nearest to the class center were
398 sampled as members of the final dataset. The final dataset comprised $12 \times 24,000$ simulated
399 cryo-EM particle images, labeled with pose angle class and structure type.

400

401 **Datasets for CryoRanker evaluation.**

402 We downloaded four types of genuine cryo-EM particle images from the EMPIAR
403 database for CryoRanker evaluation. For each type, movies or micrographs were
404 downloaded from EMPIAR and imported into CryoSPARC. Subsequently, we performed
405 motion correction, CTF estimation, particle picking, and particle clustering using
406 CryoSPARC, same as described in previously. For each type, we randomly selected 80,000
407 particle images after particle picking. The quality of the particles was scored manually in
408 the same manner as described before (Extended Data Fig. 1).

409

410 **Datasets for CryoSolver evaluation**

411 To evaluate the ability of Cryo-IEF in classifying heterogeneous structures, we constructed
412 a dataset containing particles from four different structures. The PDB IDs are 8C9C
413 (containing 9,076 particles, 12.75%), 8C99 (containing 44,366 particles, 62.32%), 8C93
414 (containing 11,975 particles, 16.82%), and 8C93 (containing 5,772 particles, 8.11%). This
415 dataset is a reduced version of the Ribosemby dataset in CryoBench⁴³. The original
416 Ribosemby dataset contains 16 types of structures. We resampled the dataset to include
417 only four types and excluded the other 12 types with high similarity to these four structures.

418

419 **Datasets for CryoWizard evaluation**

420 For the evaluation of the automated pipeline (CryoWizard), we downloaded six datasets
421 from the EMPIAR database: EMPIAR-10556, EMPIAR-10292, EMPIAR-10405,
422 EMPIAR-10250, EMPIAR-10876, and EMPIAR-10217 (Table 2). EMPIAR-10217 and
423 EMPIAR-10096 contain particles with preferred orientation problem, which were used to
424 evaluate the pipeline's performance in addressing this issue.

425

426 **Evaluation metrics**

427 **k-NN score**

428 We utilize the k-NN score to assess the performance of cryo-IEF in distinguishing particles
429 based on different structures, pose angles, and quality. For each data point in the test
430 dataset, the k-NN algorithm predicts the class by examining the k nearest neighbors and
431 taking a majority vote (in our experiments, k is set to 1). The predicted class for the i -th test
432 data point is \hat{y}_i , and the actual class is y_i . The score is calculated as follows:

$$433 \quad Score = \frac{1}{N} \sum_{i=1}^N \mathbf{1}(\hat{y}_i = y_i), \quad (1)$$

434 where $\mathbf{1}(\hat{y}_i = y_i)$ is an indicator function that equals 1 if $\hat{y}_i = y_i$ (i.e., the prediction is
435 correct), and 0 otherwise.

436

437 **AUC and AP**

438 AUC (Area Under the Curve) and AP (Average Precision) are two widely used metrics for
439 evaluating the performance of binary classification models. In our experiments, we employ
440 AUC and AP to evaluate the performance of the CryoRanker in distinguishing good

441 particles from junk particles. AUC represents the area under the receiver operating
442 characteristic (ROC) curve and can be computed as:

443
$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR}), d(\text{FPR}), \quad (2)$$

444 where TPR is the true positive rate and FPR is the false positive rate. AP measures the
445 average precision at different levels of recall and is calculated as:

446
$$\text{AP} = \sum_{n=1}^N (\text{Recall}_n - \text{Recall}_{n-1}) \times \text{Precision}_n, \quad (3)$$

447

448 where N is the number of thresholds, Recall_n and Precision_n are the recall and precision at
449 the n -th threshold.

450

451 Spearman and Pearson score

452 The Spearman and Pearson scores evaluate the correlation between predicted scores and
453 labeled scores. The Pearson correlation coefficient (often denoted as r) is calculated as:

454
$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (4)$$

455 where \bar{x} is the mean of the predicted scores x . \bar{y} is the mean of the labeled scores y .

456 Spearman score can be calculated as:

457
$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}, \quad (5)$$

458 where d_i is the difference between the predicted score and the labeled score of the i -th data
459 point.

460

461 **Pre-training of Cryo-IEF**

462 To enhance the training strategy for cryo-EM image data and improve stability, we
463 implemented several modifications while generally adhering to the MOCO v3 framework.

464

465 **Optimizer**

466 We use the AdamW optimizer with a batch size of 2048, which empirically demonstrated
467 good performance. The learning rate is set as $\text{lr} \times \text{BatchSize}/256$, with the base learning
468 rate lr set to 1×10^{-5} . A learning rate warmup is applied during the initial 5 epochs, after
469 which the learning rate is decayed using a cosine annealing schedule.

470

471 **Pre-training loss**

472 The contrastive loss function is defined as in Equation 6, where the temperature parameter
473 τ is set to 0.5.

$$474 \quad \mathcal{L}_q = -\log \frac{\exp(q \cdot k^+/\tau)}{\exp(q \cdot k^+/\tau) + \sum k^- \exp(q \cdot k^-/\tau)}. \quad (6)$$

475

476 **Data augmentation**

477 We applied random cropping, color jittering, blurring, and solarization to the input images.
478 Random rotation was also utilized to ensure the model's rotation invariance. Before
479 inputting the images into the model, they were resized to 224×224 pixels and normalized
480 by subtracting the dataset mean and dividing by the standard deviation.

481

482 **Model architecture**

483 The Vision Transformer (ViT) architecture⁵⁴ is used as the backbone of Cryo-IEF. In this
484 paper, experiments are conducted using the ViT-B/14 backbone by default, which consists
485 of a 12-layer transformer encoder with a hidden size of 768 and 12 attention heads.
486 The input patch size is 14×14, which empirically demonstrated good performance.
487 Following MOCO v3, we replaced the Layer Normalization (LN) layer with a Batch
488 Normalization (BN) layer in the ViT backbone's MLP blocks (Fig. 1).
489 For cryo-IEF, the two encoders f_q and f_k have the same backbone and projection head
490 architecture, with f_q having an additional prediction head. The projection head is a 3-layer
491 MLP, and the prediction head is a 2-layer MLP (Extended Data Fig. 6a,b). Both have a
492 hidden size of 4096 and an output layer size of 256. All MLPs include a Batch
493 Normalization (BN) layer before the ReLU activation function. During training, f_k is
494 updated with momentum as $f_k = m f_k + (1 - m) f_q$, where m is set to 0.99, in alignment
495 with MOCO v3.

496

497 **Scale-up test**

498 In this section, we assessed the impact of dataset size and model backbone parameters
499 (Extended Data Fig. 6d,e). By randomly selecting different particle datasets from the
500 original dataset, we created three additional datasets containing 25%, 50%, and 75% of the
501 particle types, in addition to the original dataset (100% particle types). As the dataset size
502 increased, the performance of cryo-IEF consistently improved. We also pre-trained cryo-
503 IEF with varying backbone sizes. Notably, larger backbones did not yield improved
504 performance, likely due to insufficient training data hindering the effective training of

505 larger models. Thus, we posit that further increased datasets may result in additional
506 performance gains.

507

508 **CryoRanker: fine-tuning cryo-IEF with labeled datasets**

509 **Training strategy**

510 During the fine-tuning process, we applied a combination of cross-entropy and margin
511 ranking loss functions to train the model. Let $\mathcal{D} = (x_i, y_i)_{i=1}^N$ represent the dataset, where
512 x_i is an input image and y_i is the corresponding label, constrained such that $y_i \in [0,1]$. The
513 model outputs a two-dimensional vector $p_i = [p_{i,0}, p_{i,1}]$ for each input image x_i , where:

514

$$p_{i,0} = \frac{e^{z_{i,0}}}{e^{z_{i,0}} + e^{z_{i,1}}}, \quad p_{i,1} = \frac{e^{z_{i,1}}}{e^{z_{i,0}} + e^{z_{i,1}}}. \quad (7)$$

515

516 Here, p_i is derived from applying the softmax function to the output logits $[z_{i,0}, z_{i,1}]$. The
517 predicted score \hat{y}_i of the image is defined as the probability in the first dimension: $p_{i,1}$.
518 The Margin Ranking Loss is applied to the predicted scores to enforce a ranking between
519 pairs of images. For pairs (x_i, x_j) where $y_i > y_j$:

520

$$\mathcal{L}_{\text{MR}} = \sum_{i,j} \max(0, \Delta - (\hat{y}_i - \hat{y}_j)), \quad (8)$$

521 where Δ is a margin hyperparameter, empirically set to 0.2 in our experiments.
522 For images with labels y_i greater than α , we label them as 1, and for those with labels less
523 than or equal to $1 - \alpha$, we label them as 0:

524

$$\tilde{y}_i = \begin{cases} 1, & \text{if } y_i > \alpha \\ 0, & \text{if } y_i \leq 1 - \alpha \end{cases}, \quad (9)$$

525 In our experiments, we empirically set α to 0.7. The Cross-Entropy Loss is then applied to
526 these labeled elements:

527
$$\mathcal{L}_{\text{CE}} = - \sum_{i \in \mathcal{I}} [\tilde{y}_i \log(p_{i,1}) + (1 - \tilde{y}_i) \log(p_{i,0})], \quad (10)$$

528 where \mathcal{I} is the set of indices i such that $y_i > \alpha$ or $y_i \leq 1 - \alpha$.

529 The final loss is the sum of the Margin Ranking Loss and the Cross-Entropy Loss:

530
$$\mathcal{L} = \mathcal{L}_{\text{MR}} + \mathcal{L}_{\text{CE}}. \quad (11)$$

531

532 **Ablation test**

533 We evaluated Cryo-IEF fine-tuned with different parameters and discovered that fine-
534 tuning the entire network yielded the best performance (Extended Data Fig. 6f), aligning
535 with findings from previous studies in traditional computer vision^{28-30,55}. We also found that
536 the model fine-tuned with both loss functions outperformed those using either loss function
537 alone in distinguishing particles of varying quality (Extended Data Fig. 6g). Therefore, for
538 our experiments, we utilized CryoRanker fine-tuned with the full network and combined
539 loss functions by default.

540

541 **Implementation of CryoSolver**

542 Our heterogeneous structural reconstruction experiments are conducted using the DRGN-
543 AI framework⁴². Compared to other heterogeneous structural reconstruction methods^{21,40,41},
544 DRGN-AI offers broader application scenarios due to its superior ab initio reconstruction
545 performance. In our experiments, we replaced the encoder of DRGN-AI with the pre-
546 trained Cryo-IEF model, while keeping its decoder unchanged. The new model was named

547 CryoSolver. In CryoSolver, the decoder consists of three hidden residual layers, each with
548 256 units and ReLU activations. The optimization settings, including hierarchical pose
549 search for 500,000 particles and stochastic gradient descent for 100 epochs, remain the
550 same as in the original DRGN-AI framework. In principle, Cryo-IEF can be easily
551 integrated into other heterogeneous structural reconstruction methods.

552

553 **Implementation of CryoWizard**

554 After completing the model training, quality score for input particles can be automatically
555 assessed. However, the optimal number of particles for reconstruction is not known a priori.
556 To address this, we have developed an automated algorithm for selecting the optimal
557 numbers of top ranked particles based on the trained model (Extended Data Fig. 3).

558

559 **Predicting particle scores with CryoRanker**

560 The first stage of CryoWizard involves predicting the scores of particles using CryoRanker.
561 The pipeline program calls CryoRanker to evaluate and record the scores of all obtained
562 particles. Subsequently, the particles are ranked from highest to lowest scores. Assume we
563 have a set of extracted particles $P = p_1, p_2, \dots, p_m$, where each particle p_i is associated
564 with a score s_i . The sorted set of particles is $P_{\text{sorted}} = p_1, p_2, \dots, p_m$, ordered such that $s_1 \geq$
565 $s_2 \geq \dots \geq s_m$. During this step, extracted features from particle images are saved locally as
566 they may be required in subsequent steps to address the preferred orientation problem.

567

568 **Optimal subset search for initial model reconstruction**

569 The second stage of CryoWizard involves searching for the optimal subset of particles for
570 initial model reconstruction. Top-ranked scores are selected during this process. In the
571 pipeline, two parallel searches are conducted with particle numbers of 50,000 and 100,000,
572 individually. The resolution of the refined structures is evaluated, and preferred orientation
573 analysis is performed for each structure. If both structures have a cFAR⁵⁶ value greater than
574 0.15, the one with the best resolution is used as the initial structure for the following Non-
575 uniform refinement. If neither structure meets the cFAR threshold of 0.15, an additional
576 module for addressing the preferred orientation problem will be initiated (Fig. 6a and
577 Extended Data Fig. 3). Further details are discussed in the section after next. In our
578 experiments, initial model reconstruction, refinement, and preferred orientation analysis are
579 conducted using the Ab-Initio Reconstruction, Non-uniform Refinement, and Orientation
580 Diagnostics functions in CryoSPARC¹⁴, with all settings maintained at their default values.
581

582 Optimal subset search for resolving the final structure

583 The third stage of CryoWizard involves searching for the optimal number of particles for
584 the final structural refinement. This step is analogous to the second step which is also based
585 on the ranked scores of the particles. Let k represent the number of particles with scores
586 greater than τ (where τ is empirically set to 0.4):

$$587 \quad k = \max\{j \mid s_j > \tau\}. \quad (12)$$

588 Thus, the subset of particles with scores exceeding τ is $P_{\text{threshold}} = \{p_1, p_2, \dots, p_k\}$, where
589 the particles in $P_{\text{threshold}}$ is sorted by scores. $P_{\text{threshold}}$ can be divided into n equal parts, with
590 the i -th part denoted as $P_i = \{p_{(i-1)\cdot k' + 1}, \dots, p_{i\cdot k'}\}$, where $k' = \left\lfloor \frac{k}{n} \right\rfloor$. Each refinement

591 utilizes the top j parts of particles, represented as $\cup_{i=1}^j P_i$, where $j \in \{0,1,2, \dots, n\}$.
592 Refinement is performed based on the initial model obtained from the previous step. For all
593 structures that meet the cFAR threshold of 0.15, the structure with the best resolution is
594 selected as the final reconstruction. In our experiments, n is set to 4, as larger values are
595 computationally expensive and yield minimal improvements. Refinement is conducted
596 using the Non-uniform Refinement function provided by CryoSPARC⁵⁷, with all settings
597 maintained at their default values.

598

599 **Implementation details of the module for addressing preferred orientation issues**
600 The normal pipeline of CryoWizard is marked with blue arrows, while the additional
601 pipeline for addressing the preferred orientation problem is marked with orange arrows
602 (Fig. 6a and Extended Data Fig. 3). The preferred orientation problem is diagnosed by
603 calculating the cFAR value of the refined structure during the initial model search step. If
604 all searched structures have a cFAR value lower than 0.15, the preferred orientation
605 treatment pipeline is initiated. Particles with higher scores are first selected and clustered
606 into n' groups (n' is set to 8 in our experiments) using k-means++ in feature space (with
607 the feature data saved locally in advance). The particles in each group are then selected for
608 initial model reconstruction and refinement. For all structures with a cFAR value higher
609 than 0.15, the one with the best resolution is used as the initial structure for the next round
610 of refinement.

611

612 **Integration of existing tools**

613 Given that functions such as motion correction, CTF estimation, particle picking, and
614 extraction can now be performed automatically, we have integrated these functions into our
615 pipeline program. This integration enables a fully automated end-to-end process, from the
616 movies/micrographs to the final optimized refined structure. In the diagram of the full
617 pipeline of CryoWizard, the steps marked with dashed borders indicate these existing tools
618 that have been incorporated into the pipeline (Extended Data Fig. 3). In our
619 implementation, we use CryoSPARC (version 4.6.0) as the primary platform for these
620 functions, and the pipeline program is implemented in Python by interfacing with
621 cryosparc-tools, an open-source Python library that enables scripting access to the
622 CryoSPARC cryo-EM software package. Theoretically, this pipeline program can be used
623 with any cryo-EM software that provides equivalent functions.

624

625 **Computational resources**

626 For pre-training, we utilized four NVIDIA A100 GPUs, each with 80 GB of memory, for a
627 total duration of 113.5 hours. For fine-tuning, we employed two NVIDIA A40 GPUs, each
628 with 40 GB of memory, for 28.5 hours. In the heterogeneous structural reconstruction
629 experiments, we used four NVIDIA A100 GPUs, each with 40 GB of memory, for 6.7
630 hours, processing a total of 71,189 particles. For the inference of CryoRanker, we utilized
631 four NVIDIA V100 GPUs, each with 32 GB of memory, achieving a performance of 35
632 minutes per million particles.

633

634

635 **Data availability**

636 Cryo-EM micrographs data and particle image data from EMPIAR are available at
637 <https://www.ebi.ac.uk/empiar/>. Density maps used to generate simulated cryo-EM
638 particle images are downloaded from EMDB, which is available at <https://www.ebi.ac.uk/emdb/>. Particles data from cryoPPP are available at https://calla.rnet.missouri.edu/cryo_ppp/. The raw data for 12 simulated and 4 genuine particle datasets can be found at
641 https://drive.google.com/drive/folders/1wSdOlCdUD3LX2UJYiQ3BNJfmaNSm9zV_?usp=sharing and https://drive.google.com/drive/folders/1KxR7ZDeH_6dkyCzoHow2ANKAeGgfyTxd?usp=sharing, respectively. The resampled version of the CryoBench
643 Ribosemby datasetis available at <https://drive.google.com/drive/folders/1AXZ2hQwYowfQczRIDpSNjpmwqpKwK4PQ?usp=sharing>.
645 The reconstructed results obtained from CryoSolver and CryoWizard can be accessed
646 at https://drive.google.com/drive/folders/1EKv5atcQA7n-VSTQ2_DW-qoVGr6g4C_X?usp=sharing and <https://drive.google.com/drive/folders/1cr8nxCvkq5EAE17uZYGrJnYaxuKTwegr?usp=sharing>, respectively.
649
650

651 **Code availability**

652 The codes with introduction details are available at <https://github.com/westlake-repl/Cryo-IEF>, which is based on PyTorch.
653
654

655 **References**

- 656
- 657 1 Nogales, E. The development of cryo-EM into a mainstream structural biology
658 technique. *Nature methods* **13**, 24-27 (2016).
- 659 2 Cheng, Y. Single-particle cryo-EM—How did it get here and where will it go.
660 *Science* **361**, 876-880 (2018).
- 661 3 Bai, X.-C., McMullan, G. & Scheres, S. H. How cryo-EM is revolutionizing
662 structural biology. *Trends in biochemical sciences* **40**, 49-57 (2015).
- 663 4 Frank, J. Advances in the field of single-particle cryo-electron microscopy over the
664 last decade. *Nature protocols* **12**, 209-212 (2017).
- 665 5 Holcomb, J. *et al.* Protein crystallization: Eluding the bottleneck of X-ray
666 crystallography. *AIMS biophysics* **4**, 557 (2017).
- 667 6 Scheiner, G. The Resolution Revolution. *Diabetes Self Manag* **32**, 28-29, 33 (2015).
- 668 7 Amunts, A. *et al.* Structure of the yeast mitochondrial large ribosomal subunit.
669 *Science* **343**, 1485-1489 (2014).
- 670 8 Liao, M., Cao, E., Julius, D. & Cheng, Y. Structure of the TRPV1 ion channel
671 determined by electron cryo-microscopy. *Nature* **504**, 107-112 (2013).
- 672 9 McMullan, G., Faruqi, A. & Henderson, R. Direct electron detectors. *Method Enzymol* **579**, 1-17 (2016).
- 673 10 Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for
674 improved cryo-electron microscopy. *Nature methods* **14**, 331-332 (2017).
- 675 11 Li, X. *et al.* Electron counting and beam-induced motion correction enable near-
676 atomic-resolution single-particle cryo-EM. *Nature methods* **10**, 584-590 (2013).
- 677 12 Bai, X.-c., Fernandez, I. S., McMullan, G. & Scheres, S. H. Ribosome structures to
678 near-atomic resolution from thirty thousand cryo-EM particles. *elife* **2**, e00461
679 (2013).
- 680 13 Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM
681 structure determination. *Journal of structural biology* **180**, 519-530 (2012).
- 682 14 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC:
683 algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods*
684 **14**, 290-296, doi:10.1038/nmeth.4169 (2017).
- 685 15 Zhou, Y., Moscovich, A., Bendory, T. & Bartesaghi, A. Unsupervised particle
686 sorting for high-resolution single-particle cryo-EM. *Inverse Problems* **36**, 044002
687 (2020).
- 688 16 Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure
689 determination in RELION-3. *Elife* **7**, doi:10.7554/elife.42166 (2018).
- 690 17 Bepler, T. *et al.* Positive-unlabeled convolutional neural networks for particle
691 picking in cryo-electron micrographs. *Nat Methods* **16**, 1153-1160,
692 doi:10.1038/s41592-019-0575-8 (2019).
- 693 18 Wagner, T. *et al.* SPHIRE-crYOLO is a fast and accurate fully automated particle
694 picker for cryo-EM. *Commun Biol* **2**, 218, doi:10.1038/s42003-019-0437-z (2019).
- 695 19 Kimanis, D., Dong, L., Sharov, G., Nakane, T. & Scheres, S. H. W. New tools for
696 automated cryo-EM single-particle analysis in RELION-4.0. *Biochem J* **478**, 4169-
697 4185, doi:10.1042/BCJ20210708 (2021).
- 698

- 699 20 Li, Y., Cash, J. N., Tesmer, J. J. G. & Cianfrocco, M. A. High-Throughput Cryo-
700 EM Enabled by User-Free Preprocessing Routines. *Structure* **28**, 858-869 e853,
701 doi:10.1016/j.str.2020.03.008 (2020).
- 702 21 Zhong, E. D., Bepler, T., Berger, B. & Davis, J. H. CryoDRGN: reconstruction of
703 heterogeneous cryo-EM structures using neural networks. *Nat Methods* **18**, 176-
704 185, doi:10.1038/s41592-020-01049-4 (2021).
- 705 22 Pfab, J., Phan, N. M. & Si, D. DeepTracer for fast de novo cryo-EM protein
706 structure modeling and special studies on CoV-related complexes. *Proc Natl Acad
707 Sci U S A* **118**, doi:10.1073/pnas.2017525118 (2021).
- 708 23 Jamali, K. et al. Automated model building and protein identification in cryo-EM
709 maps. *Nature* **628**, 450-457, doi:10.1038/s41586-024-07215-4 (2024).
- 710 24 Scheres, S. H. Processing of structurally heterogeneous cryo-EM data in RELION.
711 *Method Enzymol* **579**, 125-157 (2016).
- 712 25 Zhu, D. et al. Correction of preferred orientation-induced distortion in cryo-
713 electron microscopy maps. *Science Advances* **10**, eadn0092 (2024).
- 714 26 Zhang, H. et al. Addressing preferred orientation in single-particle cryo-EM through
715 AI-generated auxiliary particles. *bioRxiv*, 2023.2009. 2026.559492 (2023).
- 716 27 Liu, Y., Fan, H., Hu, J. & Zhou, Z. H. Overcoming the preferred orientation
717 problem in cryoEM with self-supervised deep-learning. *bioRxiv*, 2024.2004.
718 2011.588921 (2024).
- 719 28 Chen, T., Kornblith, S., Norouzi, M. & Hinton, G. in *International conference on
720 machine learning*. 1597-1607 (PMLR).
- 721 29 He, K., Fan, H., Wu, Y., Xie, S. & Girshick, R. in *Proceedings of the IEEE/CVF
722 conference on computer vision and pattern recognition*. 9729-9738.
- 723 30 Chen, X., Xie, S. & He, K. in *Proceedings of the IEEE/CVF international
724 conference on computer vision*. 9640-9649.
- 725 31 Oquab, M. et al. Dinov2: Learning robust visual features without supervision. *arXiv
726 preprint arXiv:2304.07193* (2023).
- 727 32 Zhou, Y. et al. A foundation model for generalizable disease detection from retinal
728 images. *Nature* **622**, 156-163, doi:10.1038/s41586-023-06555-x (2023).
- 729 33 Pai, S. et al. Foundation model for cancer imaging biomarkers. *Nat Mach Intell* **6**,
730 354-367, doi:10.1038/s42256-024-00807-9 (2024).
- 731 34 Wang, X. et al. A pathology foundation model for cancer diagnosis and prognosis
732 prediction. *Nature*, 1-9 (2024).
- 733 35 Xu, H. et al. A whole-slide foundation model for digital pathology from real-world
734 data. *Nature*, 1-8 (2024).
- 735 36 Ma, C., Tan, W., He, R. & Yan, B. Pretraining a foundation model for generalizable
736 fluorescence microscopy-based image restoration. *Nat Methods* **21**, 1558-1567,
737 doi:10.1038/s41592-024-02244-3 (2024).
- 738 37 Iudin, A. et al. EMPIAR: the Electron Microscopy Public Image Archive. *Nucleic
739 Acids Res* **51**, D1503-D1511, doi:10.1093/nar/gkac1062 (2023).
- 740 38 Dhakal, A., Gyawali, R., Wang, L. & Cheng, J. A large expert-curated cryo-EM
741 image dataset for machine learning protein particle picking. *Sci Data* **10**, 392,
742 doi:10.1038/s41597-023-02280-2 (2023).
- 743 39 El Banani, M. et al. in *Proceedings of the IEEE/CVF Conference on Computer
744 Vision and Pattern Recognition*. 21795-21806.

- 745 40 Zhong, E. D., Lerer, A., Davis, J. H. & Berger, B. in *Proceedings of the IEEE/CVF International Conference on Computer Vision.* 4066-4075.
746
747 41 Luo, Z., Ni, F., Wang, Q. & Ma, J. OPUS-DSD: deep structural disentanglement for cryo-EM single-particle analysis. *Nat Methods* **20**, 1729-1738, doi:10.1038/s41592-023-02031-6 (2023).
748
749
750 42 Levy, A. et al. Revealing biomolecular structure and motion with neural ab initio cryo-EM reconstruction. *bioRxiv*, doi:10.1101/2024.05.30.596729 (2024).
751
752 43 Jeon, M. et al. CryoBench: Diverse and challenging datasets for the heterogeneity problem in cryo-EM. *arXiv preprint arXiv:2408.05526* (2024).
753
754 44 Hu, M. et al. A particle-filter framework for robust cryo-EM 3D reconstruction. *Nature methods* **15**, 1083-1089 (2018).
755
756 45 Tan, Y. Z. et al. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat Methods* **14**, 793-796, doi:10.1038/nmeth.4347 (2017).
757
758 46 McInnes, L., Healy, J. & Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426* (2018).
759
760 47 Kuhlbrandt, W. Biochemistry. The resolution revolution. *Science* **343**, 1443-1444, doi:10.1126/science.1251652 (2014).
761
762 48 Liu, Z. et al. Determination of the ribosome structure to a resolution of 2.5 Å by single - particle cryo - EM. *Protein Science* **26**, 82-92 (2017).
763
764 49 Fan, X. et al. Single particle cryo-EM reconstruction of 52 kDa streptavidin at 3.2 Angstrom resolution. *Nat. Commun.* **10**, 2386 (2019).
765
766 50 Baxter, W. T., Grassucci, R. A., Gao, H. & Frank, J. Determination of signal-to-noise ratios and spectral SNRs in cryo-EM low-dose imaging of molecules. *Journal of structural biology* **166**, 126-132 (2009).
767
768
769 51 Palovcak, E., Asarnow, D., Campbell, M. G., Yu, Z. & Cheng, Y. Enhancing the signal-to-noise ratio and generating contrast for cryo-EM images with convolutional neural networks. *IUCrJ* **7**, 1142-1150 (2020).
770
771
772 52 Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of structural biology* **192**, 216-221 (2015).
773
774 53 Lawson, C. L. et al. EMDataBank unified data resource for 3DEM. *Nucleic acids research* **44**, D396-D403 (2016).
775
776 54 Dosovitskiy, A. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020).
777
778 55 Grill, J.-B. et al. Bootstrap your own latent-a new approach to self-supervised learning. *Advances in neural information processing systems* **33**, 21271-21284 (2020).
779
780
781 56 Baldwin, P. R. & Lyumkis, D. Non-uniformity of projection distributions attenuates resolution in Cryo-EM. *Progress in biophysics and molecular biology* **150**, 160-183 (2020).
782
783
784 57 Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat Methods* **17**, 1214-1221, doi:10.1038/s41592-020-00990-8 (2020).
785
786
787 58 Pettersen, E. F. et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein science* **30**, 70-82 (2021).
788
789
790

791 **Acknowledgments**

792 We thank the HPC Center of Westlake University for providing computational facility
793 support and technical assistance. This work was supported by the Research Center for
794 Industries of the Future (RCIF), Westlake University, the National Science Foundation of
795 China (32122042 and 32071208 to H.S., U21A20427 to F.Y.), the Ministry of Science and
796 Technology (MOST) of the People's Republic of China (2022ZD0115100 to F.Y.), the
797 Zhejiang Provincial Natural Science Foundation (DQ24C050001 to H.S.), and the Westlake
798 Education Foundation (to H.S.). We thank our colleagues Yigong Shi, Hongtao Yu, Peilong
799 Lu, Dan Ma, Qi Hu, Qiang Zhou, Jianping Wu, Zhen Yan, Zhubing Shi, and Jijie Chai for
800 generously sharing their in-house cryo-EM data. We would also like to acknowledge the
801 use of data from EMDB and EMPIAR for training our models.

802

803 **Author information**

804 These authors contributed equally: Yang Yan and Shiqi Fan.

805

806 **Authors and affiliations**

807 **Research Center for Industries of The Future (RCIF), Westlake University,**

808 **Hangzhou, Zhejiang, China**

809 Yang Yan, Shiqi Fan, Fajie Yuan & Huaizong Shen

810

811 **School of Engineering, Westlake University, Hangzhou, Zhejiang, China**

812 Yang Yan, Shiqi Fan & Fajie Yuan

813

814 **Zhejiang Key Laboratory of Structural Biology, School of Life Sciences, Westlake**

815 **University; Hangzhou, Zhejiang, China**

816 Huaizong Shen

817

818 **Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang, China**

819 Huaizong Shen

820

821 **Institute of Biology, Westlake Institute for Advanced Study, Hangzhou, Zhejiang,**

822 **China**

823 Huaizong Shen

824

825 **Contributions**

826 The project was conceived and supervised by F.Y. and H.S. Y.Y. was primarily responsible
827 for training the AI models, while S.F. mainly handled the preparation and processing of
828 Cryo-EM data as well as the construction of the automated data processing pipeline. The
829 initial draft of the manuscript was written by Y.Y. and subsequently revised and finalized
830 by F.Y. and H.S. All authors reviewed and provided feedback on the manuscript.

831

832 **Corresponding authors**

833 Correspondence to Fajie Yuan or Huaizong Shen.

834

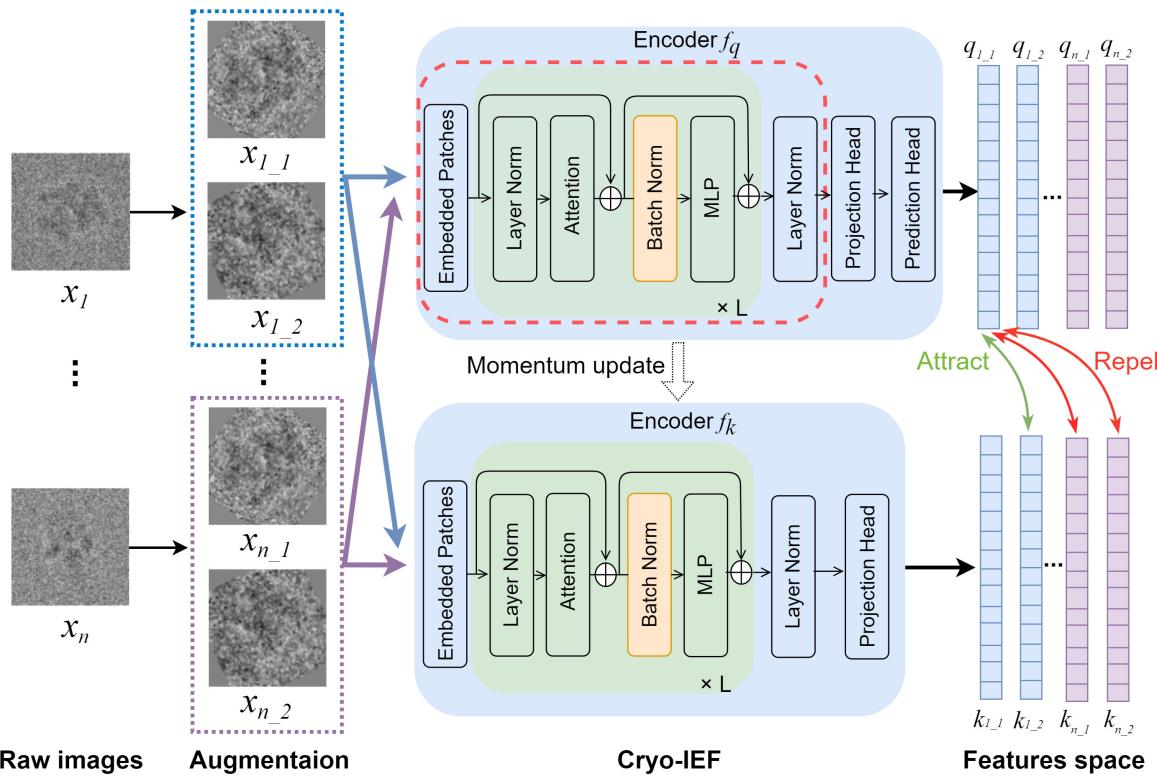
835 **Ethics declarations**

836 Competing interests

837 The authors declare no competing interests.

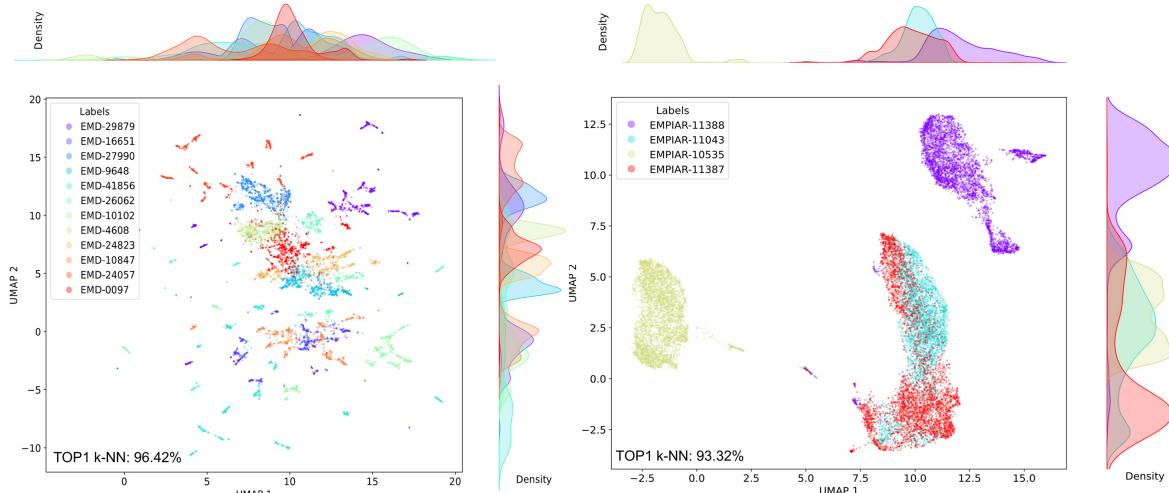
838

839 **Figures**



841 **Fig. 1 | Pre-training framework of Cryo-IEF.** Cryo-IEF utilizes a contrastive learning
842 framework for pre-training. Two distinctly augmented views are generated from each
843 particle image of the training datasets and encoded by two parallel encoders. The model is
844 trained to maximize the similarity between the two views of the same image while
845 minimizing the similarity between views from different particle images. Only the modules
846 within the red dashed box are employed for downstream cryo-EM image processing tasks.
847 For detailed information on the model architecture, please refer to the Methods.

848

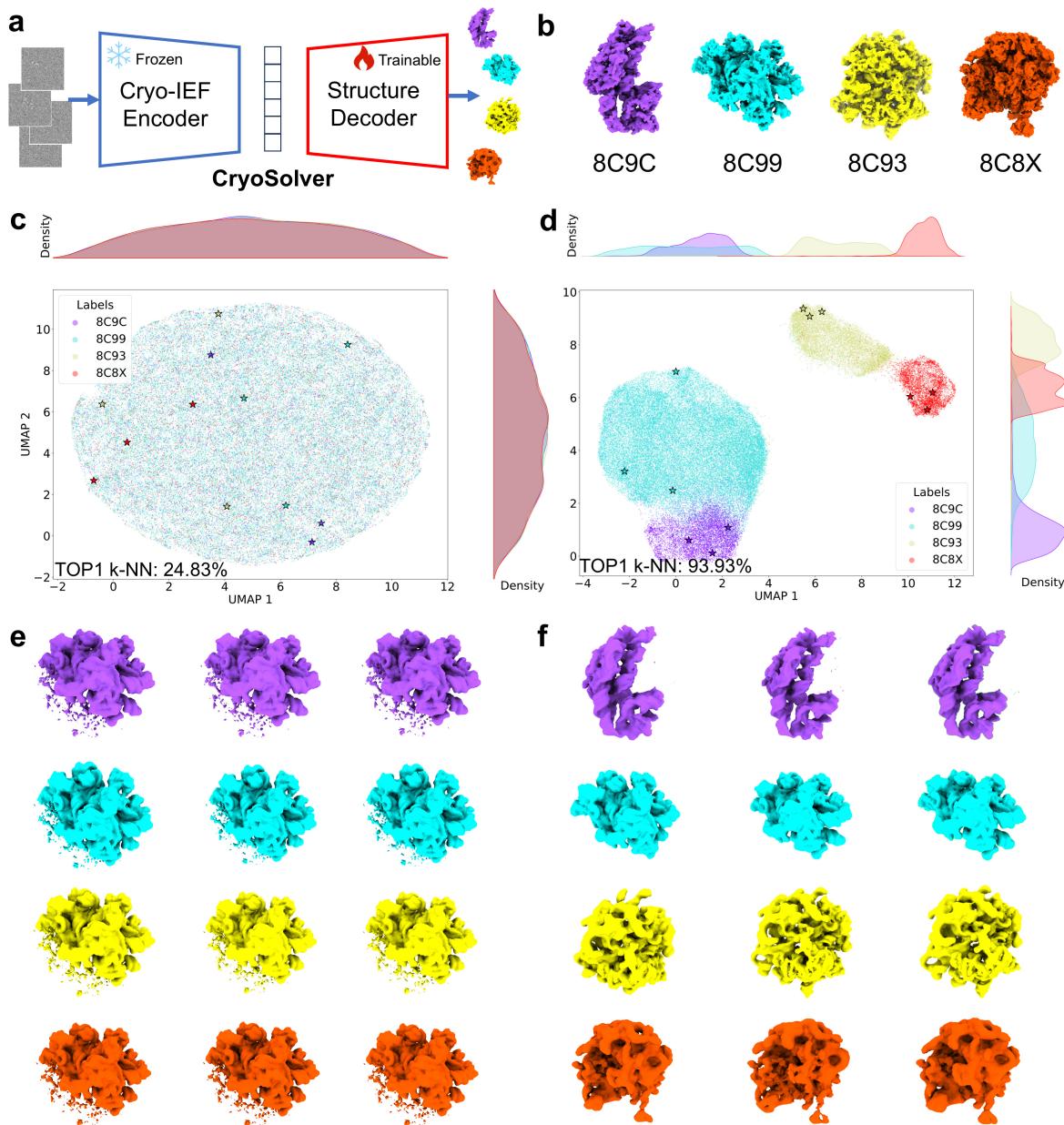


849

850 **Fig. 2 | Cryo-IEF's performance in classifying particles from different structures.**

851 Features extracted by Cryo-IEF from particles in 12 simulated datasets (left panel) and 4
852 genuine datasets (right panel) are visualized using Uniform Manifold Approximation and
853 Projection (UMAP). These particle features are color-coded and labeled with their ground-
854 truth EMDB or EMPIAR identifier numbers. The k-Nearest Neighbors (k-NN) scores for
855 the particles from the simulated and genuine datasets are 96.42% and 93.32%, respectively,
856 indicating that Cryo-IEF effectively classifies cryo-EM particles from various structures.

857

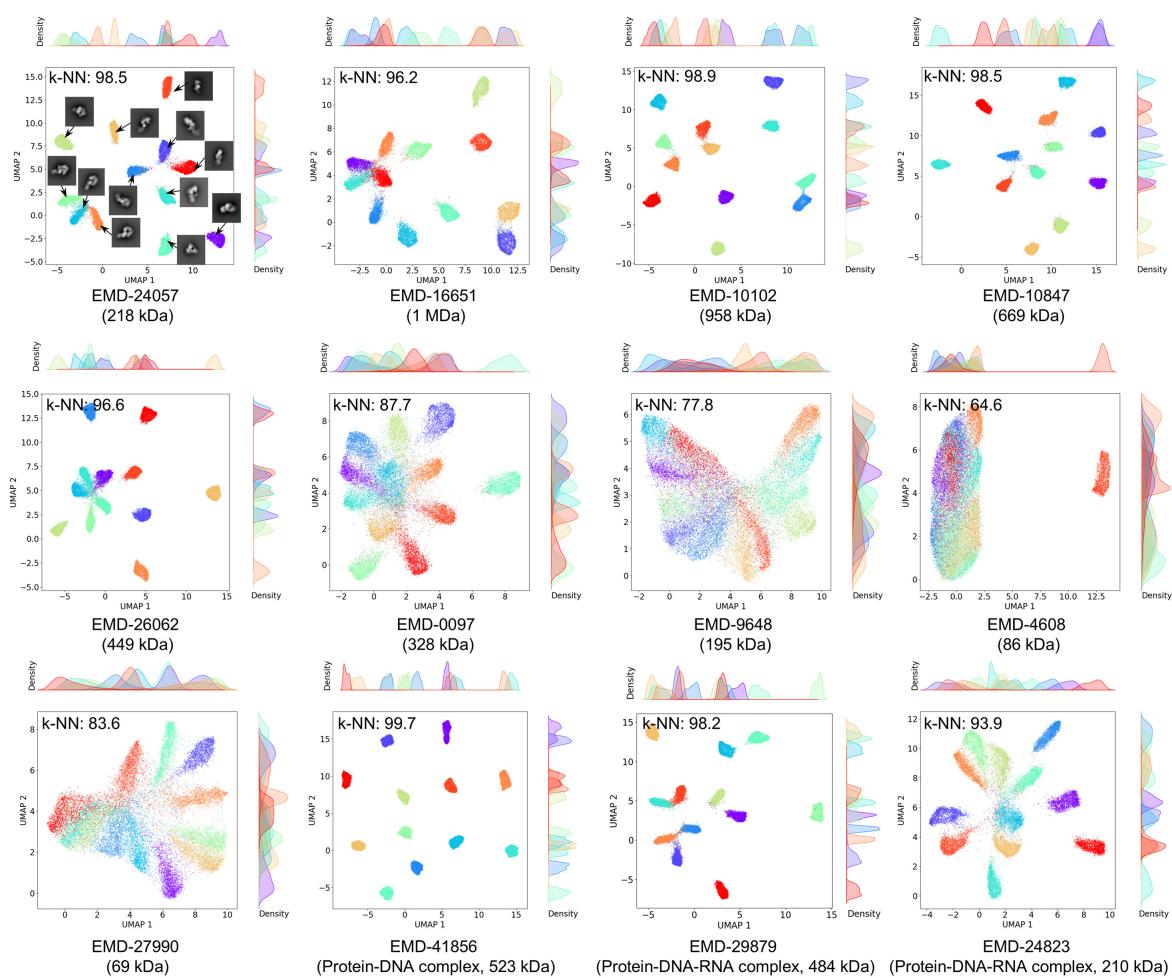


858

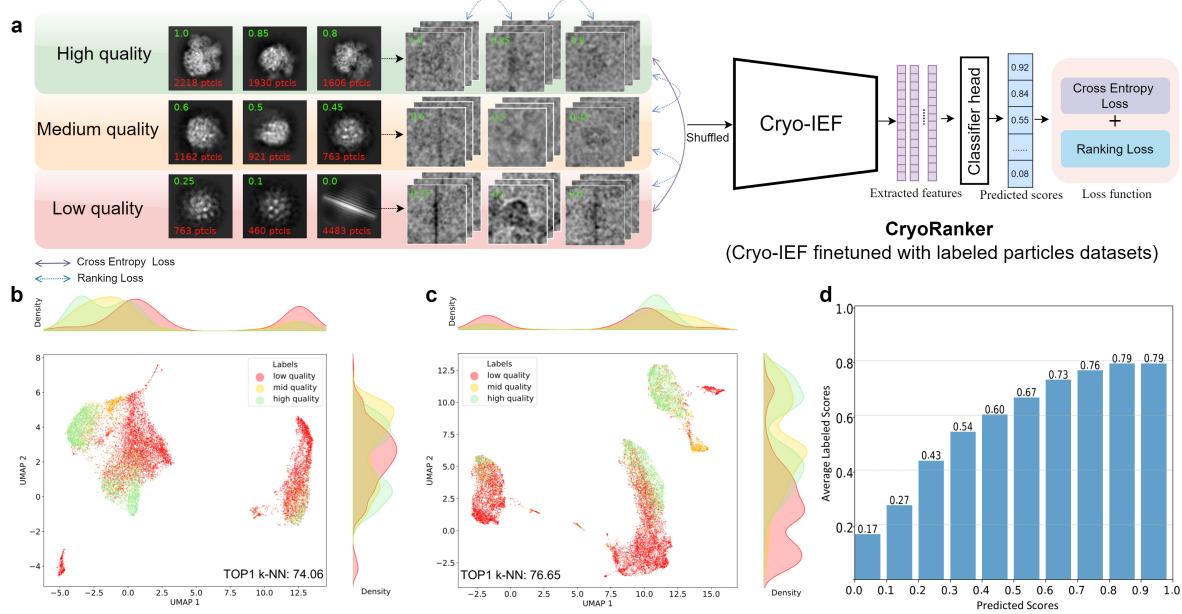
859 **Fig. 3 | Reconstruction of heterogeneous structures from features extracted by Cryo-**
860 **IEF. (a)** The framework of CryoSolver for heterogeneous structural reconstruction.
861 CryoSolver integrates Cryo-IEF with frozen parameters as the encoder and the trainable
862 structure decoder of DRGN-AI for reconstructing heterogeneous structures . **(b)** Ground-
863 truth structural maps corresponding to the test dataset are displayed, with their respective
864 PDB codes labeled. **(c)(d)** Visualization of particle features from the test dataset, color-

865 coded by their originating structures and displayed using UMAP. Panel **(c)** shows features
866 initialized randomly, while panel **(d)** shows features extracted by Cryo-IEF. **(e)(f)**
867 Reconstruction of individual structures from particle features marked with stars in panels
868 **(c)** and **(d)** using the structure decoder. The reconstructed structures are shown in
869 corresponding colors in panels **(e)** and **(f)**, respectively.

870



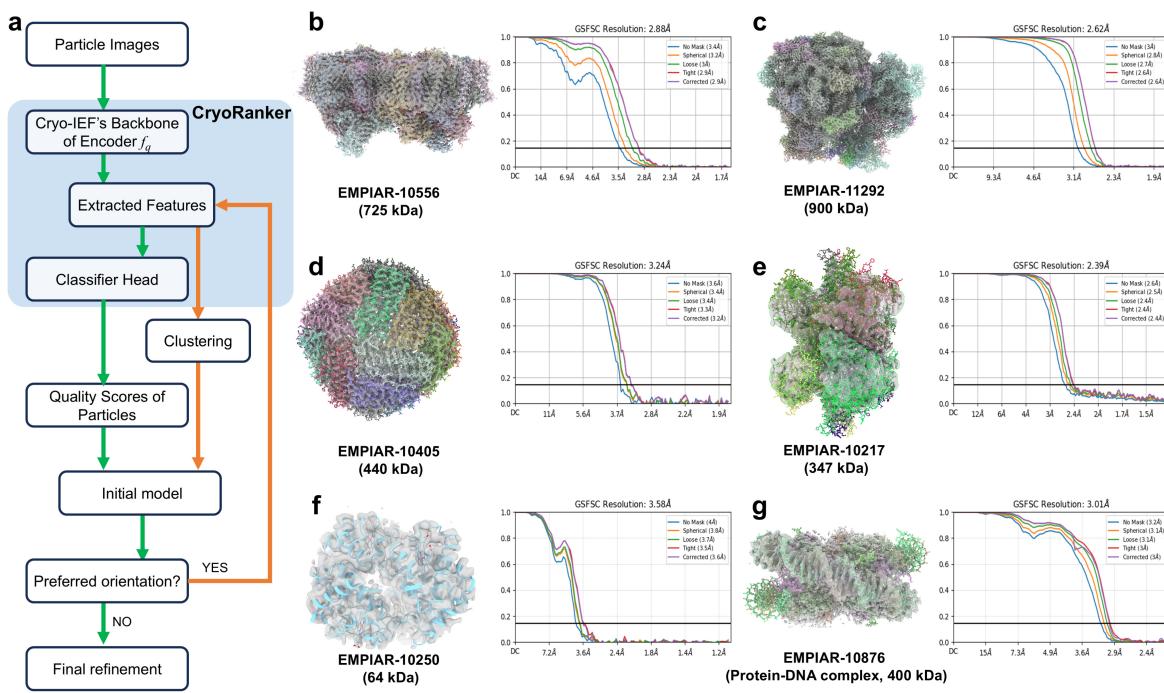
871 **Fig. 4 | Cryo-IEF's performance in clustering particles by pose.** The particle features of
872 the 12 simulated datasets, extracted by Cryo-IEF, are color-coded according to their
873 corresponding ground-truth pose clusters and visualized with UMAP. The distribution of
874 these particle features and the calculated k-NN scores demonstrate that Cryo-IEF
875 effectively clusters cryo-EM particles according to their poses. For data generated from
876 map EMD-24057, particles from each cluster were further processed using the 2D
877 classification job in CryoSPARC. The resulting 2D class averages are displayed adjacent to
878 their corresponding feature clusters in the panel.
880



881

882 **Fig. 5 | Cryo-IEF's performance in ranking particles by quality.** (a) Fine-tuning
883 framework of CryoRanker. Cryo-IEF was fine-tuned to create CryoRanker, which ranks
884 particles according to quality. During fine-tuning, particles were processed through
885 CryoSPARC's 2D classification job and manually assigned quality scores from 0 to 1,
886 reflecting the quality of their corresponding 2D class averages. The optimization employed
887 a combination of cross-entropy and margin ranking losses. Particles scoring above 0.7 were
888 deemed high-quality, those below 0.3 low-quality, and scores falling in between indicated
889 mid-quality. Further details on the quality score assignment are provided in Extended Data
890 Fig. 1. (b)(c) Distributions of features extracted by Cryo-IEF (b) and CryoRanker (c) from
891 four genuine datasets are color-coded according to their labeled qualities and visualized
892 with UMAP. The results indicate that both models can distinguish particles of different
893 qualities, with the fine-tuned CryoRanker achieving a higher k-NN score. (d) The
894 distribution of the manually labeled average quality scores for test datasets, in relation to
895 the predicted ones, reveals a positive correlation between the two sets of scores.

896

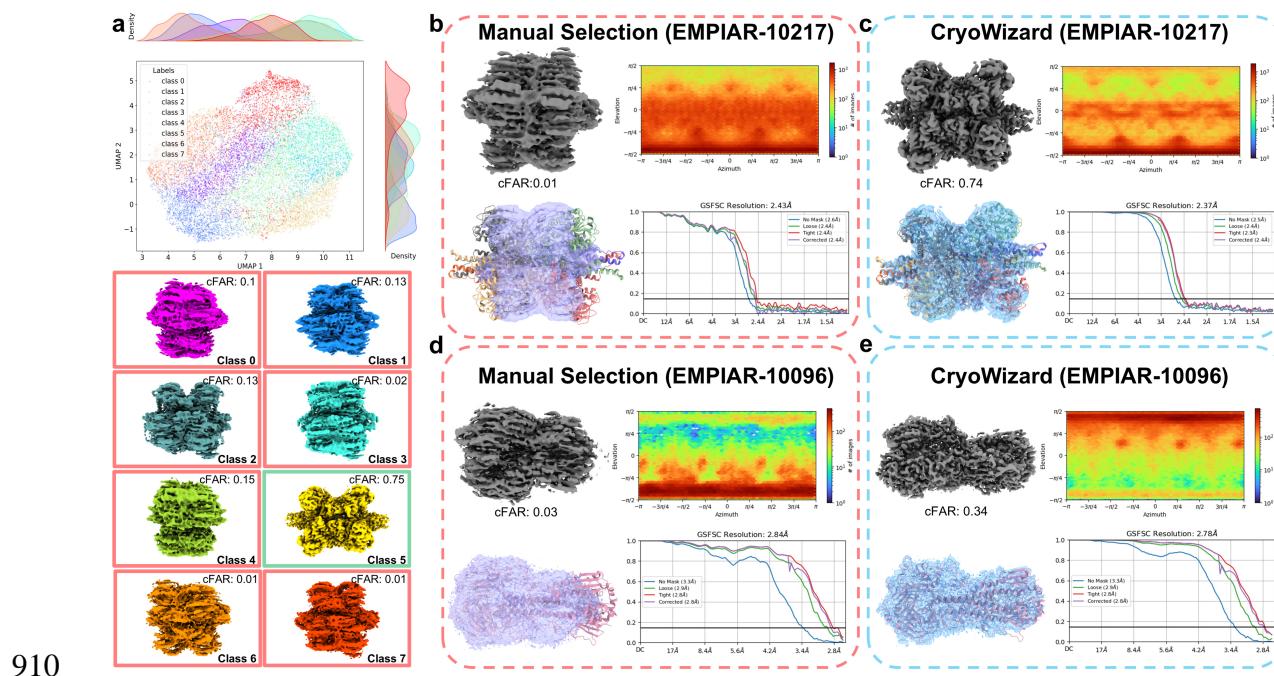


897

898 **Fig. 6 | Performance of the fully automated cryo-EM data processing pipeline,**

899 **CryoWizard.** (a) A simplified data processing flowchart of CryoWizard. CryoWizard is a
900 fully automated cryo-EM data processing pipeline with recorded cryo-EM
901 movies/micrographs or particle images as inputs and the final maps as outputs. CryoRanker
902 plays a central role in the pipeline by replacing the time- and resource-consuming, multiple
903 rounds of 2D/3D classification jobs which requires human interventions. For details, please
904 refer to the Methods and Extended Data Fig. 3. (b-g) Cryo-EM maps automatically
905 resolved by the pipeline are aligned with their respective PDB models. Fourier Shell
906 Correlation (FSC) curves for each reconstruction are displayed to indicate the resolved
907 resolutions. EMPIAR ID and molecular weights of the datasets are also labeled. The figures
908 of aligned maps and models were prepared using ChimeraX⁵⁸.

909



910 **Fig. 7 | CryoWizard effectively addresses the preferred orientation problem. (a)** For
911 particle datasets severely afflicted by the preferred orientation problem, CryoWizard
912 incorporates an additional module which clusters the Cryo-IEF-extracted particle features
913 using the K-Means++ algorithm. The clustered particles from the genuine dataset
914 EMPIAR-10217 are color-coded and visualized with UMAP. Examination of the
915 reconstructed maps from the eight classes of particles reveals that one class (Class 5) is not
916 affected by preferred orientation, providing a correct template for subsequent map
917 refinement. The conical FSC Area Ratio (cFAR) scores, which indicate the severity of the
918 preferred orientation problem, are displayed in the top right corner of each map. **(b)(c)**
919 Final refinement maps reconstructed using manually selected particles **(b)** and the
920 CryoWizard pipeline **(c)** from the EMPIAR-10217 dataset. **(d)(e)** Final refinement maps
921 reconstructed using manually selected particles **(d)** and the CryoWizard pipeline **(e)** from
922 the EMPIAR-10096 dataset.

Yan *et al*

923 the EMPIAR-10096 dataset. The details of manual selected particles for EMPIAR-10217

924 and EMPIAR-10096 datasets are shown in Extended Data Figs. 4 and 5, respectively.

925

926 **Table 1 | Test datasets of simulated cryo-EM particle images**

Particle Category	EMDB ID (Overall Molecular Weight)		
0-200 kDa	EMD-27990 (69 kDa)	EMD-4608 (86 kDa)	EMD-9648 (195 kDa)
200-600 kDa	EMD-24057 (218 kDa)	EMD-0097 (328 kDa)	EMD-26062 (449 kDa)
600-2000 kDa	EMD-10847 (669 kDa)	EMD-16651 (1 MDa)	EMD-10102 (958 kDa)
With Nucleic Acids	EMD-41856 (523 kDa)	EMD-29879 (484 kDa)	EMD-24823 (210 kDa)

927

928 **Table 2 | Test datasets of genuine cryo-EM particle images**

EMPIAR ID	Overall Molecular Weight (kDa)	Raw pixel size (Å)	Accelerating Voltage (kV)	Spherical Aberration (mm)	Total exposure dose (e/Å ²)
EMPIAR-10535 [†]	68	1.038	200	2.7	47.13
EMPIAR-11043 [†]	774	0.830	300	2.7	58.20
EMPIAR-11387 [†]	1313	0.968	200	2.7	40.00
EMPIAR-11388 ^{†*}	1500, 2200	0.800	300	4.0	40.00
EMPIAR-10556 [‡]	725	0.822	300	2.7	83.00
EMPIAR-11292 [‡]	900	0.860	300	2.7	40.00
EMPIAR-10405 [‡]	440	0.435	200	2.7	60.00
EMPIAR-10250 [‡]	64	0.556	300	2.7	69.00
EMPIAR-10876 [‡]	400	0.540	300	2.7	49.90
EMPIAR-10217 ^{‡◊}	347	0.662	300	2.7	61.25
EMPIAR-10096 [◊]	150	1.310	300	2.7	82.00

929 [†] Datasets for CryoRanker evaluation.

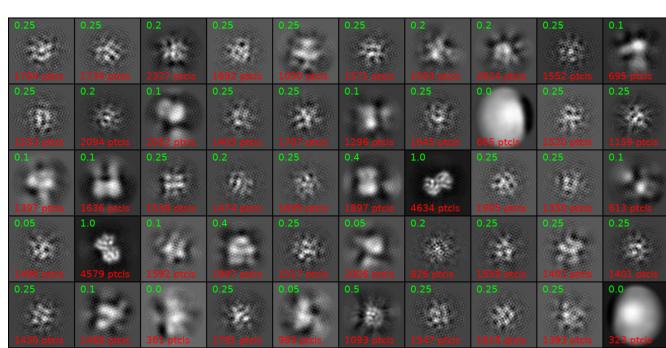
930 [‡] Datasets for CryoWizard evaluation.

931 [◊] Datasets for preferred orientation correction test.

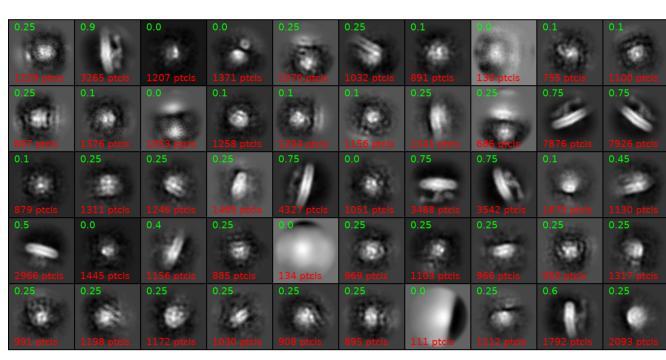
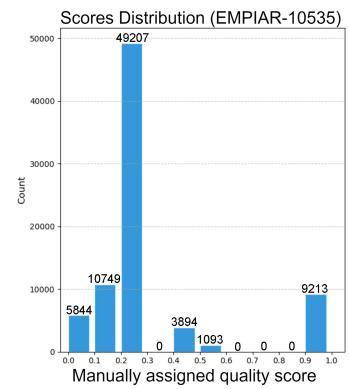
932 * This structure has multiple conformations.

933

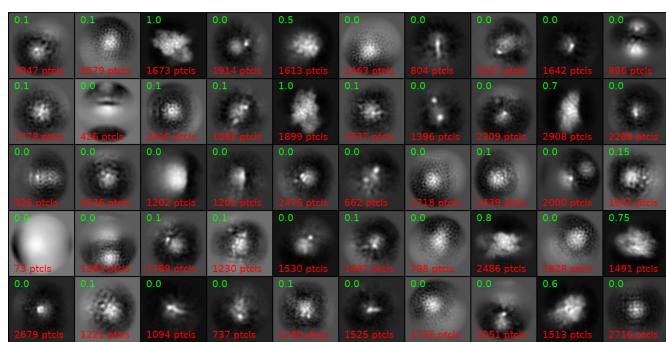
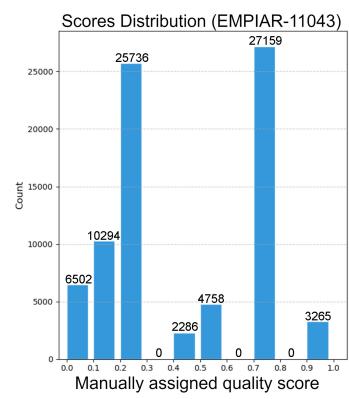
934 Extended Data Figures and Tables



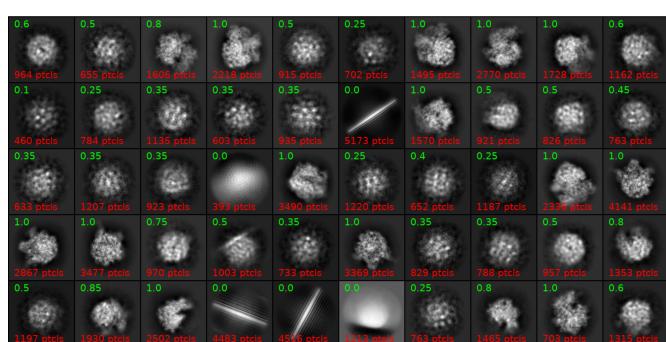
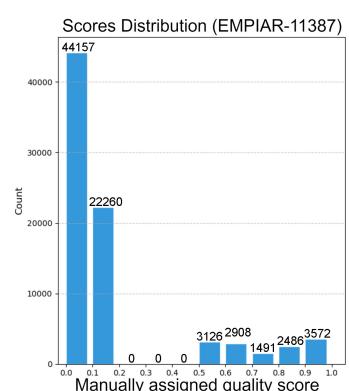
2D averages with manually assigned quality score (EMPIAR-10535)



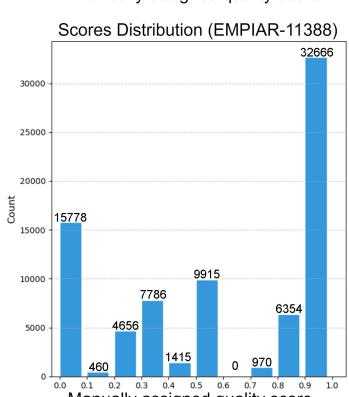
2D averages with manually assigned quality score (EMPIAR-11043)



2D averages with manually assigned quality score (EMPIAR-11387)

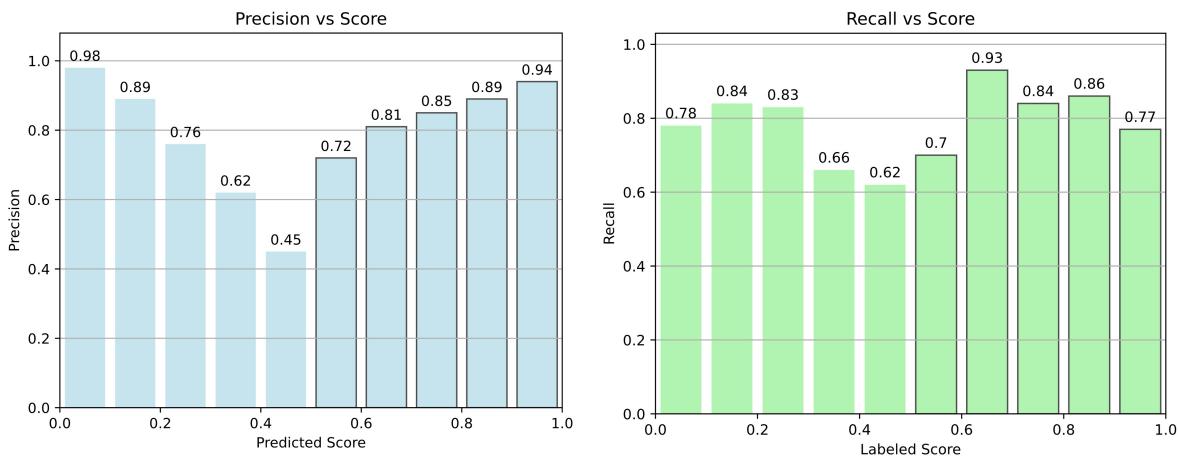


2D averages with manually assigned quality score (EMPIAR-10388)



936 **Extended Data Fig. 1 | 2D classification and quality score assignment of four genuine
937 particle datasets.** The cryo-EM particles of the four genuine datasets (EMPIAR-10535,
938 11043, 11387, and 11388) were processed using the 2D classification job in CryoSPARC
939 and scored manually based on the quality of the 2D class averages. Distributions of the
940 manually assigned particle scores for the four datasets are listed side by side with the 2D
941 classification results of their respective datasets.

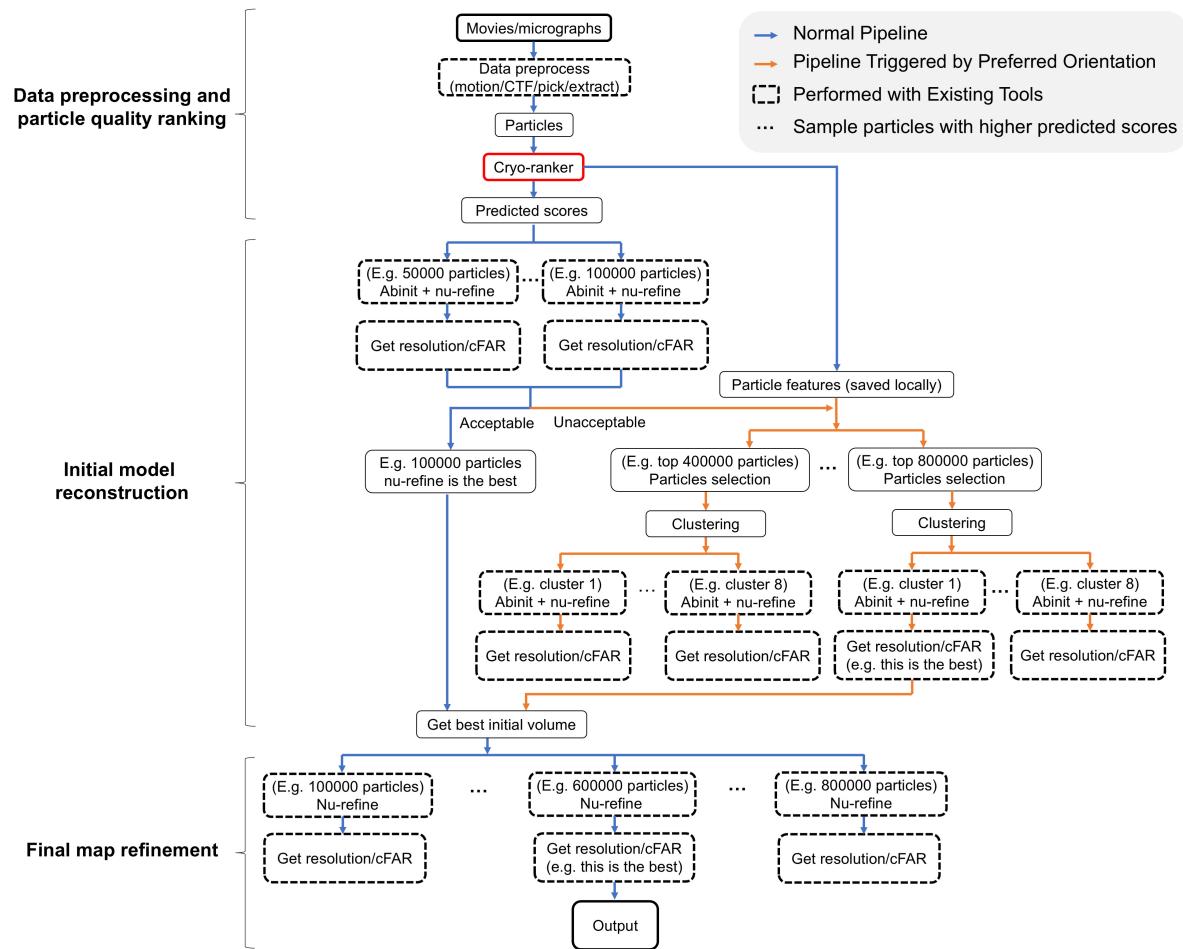
942



943

944 **Extended Data Fig. 2 | Evaluation of CryoRanker by precision and recall metrics.** The
945 precision scores in relation to the predicted particle scores (left panel) and the recall values
946 in relation to the labeled particle scores (right panel) of the four genuine particle datasets
947 are displayed, indicating the performance of the CryoRanker model.

948



949

950 **Extended Data Fig. 3 | Detailed flowchart for the fully automated cryo-EM data**

951 **processing pipeline, CryoWizard.** The default pipeline is marked by blue arrows, while
952 the pipeline for addressing the preferred orientation problem is marked by orange arrows.
953 The preferred orientation problem is diagnosed by calculating the cFAR value of the
954 refined structure during the initial model search step. Current pipeline is implemented in
955 Python and interfaces with CryoSPARC-tools, an open-source Python library that enables
956 scripted access to the CryoSPARC software package.

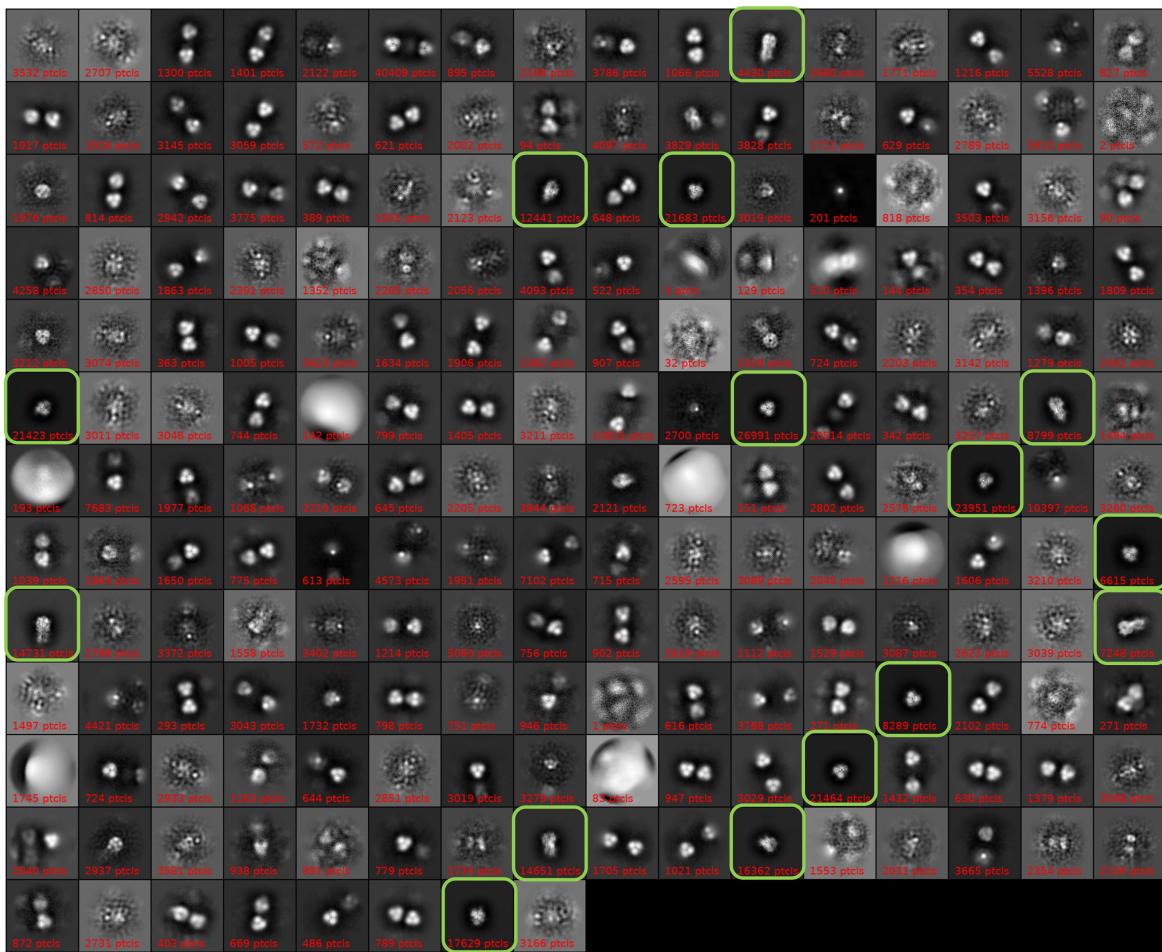
957



958

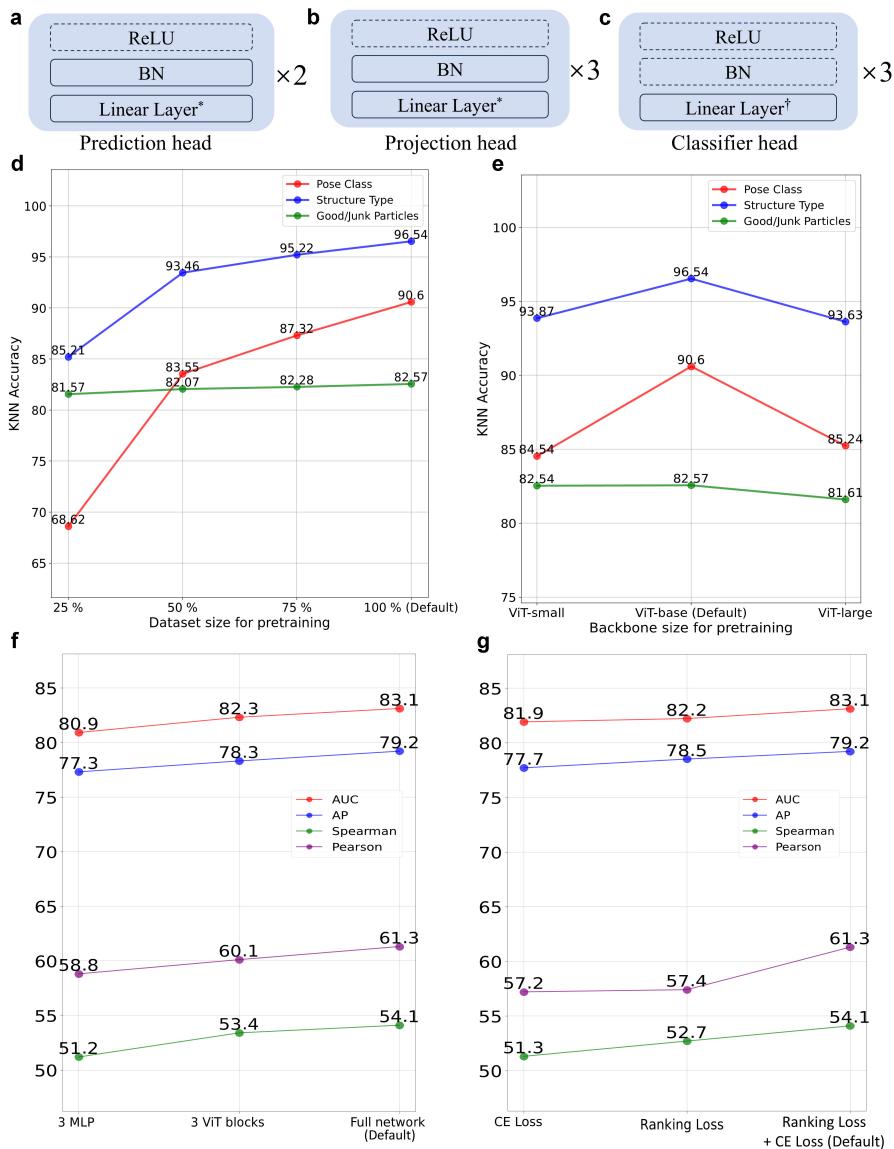
959 **Extended Data Fig. 4 | 2D classification results of particles from dataset EMPIAR-**

960 **10217.** The cryo-EM particles of the dataset EMPIAR-10217 were processed using the 2D
961 classification job in CryoSPARC. Manually selected classes for the following map
962 reconstruction are indicated by the green boxes.



964 **Extended Data Fig. 5 | 2D classification results of particles from dataset EMPIAR-**

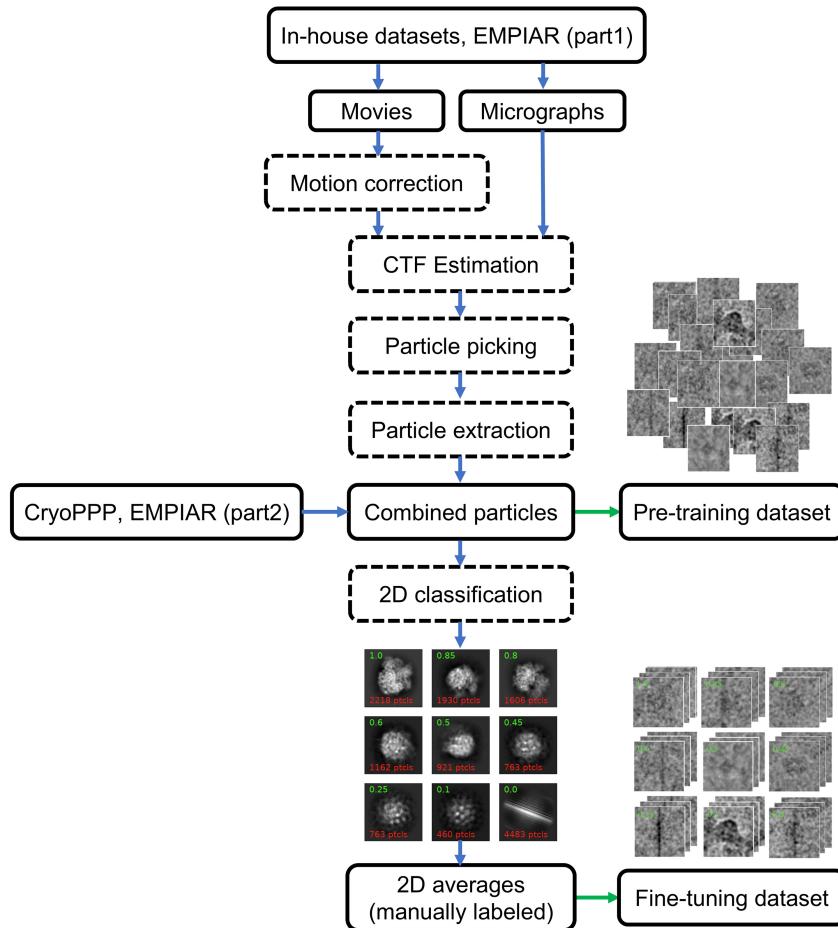
965 **10096.** The cryo-EM particles of the dataset EMPIAR-10096 were processed using the 2D
966 classification job in CryoSPARC. Manually selected classes for the following map
967 reconstruction are indicated by the green boxes.
968



969

970 Extended Data Fig. 6 | Detailed architectures and scale and ablation tests of the AI

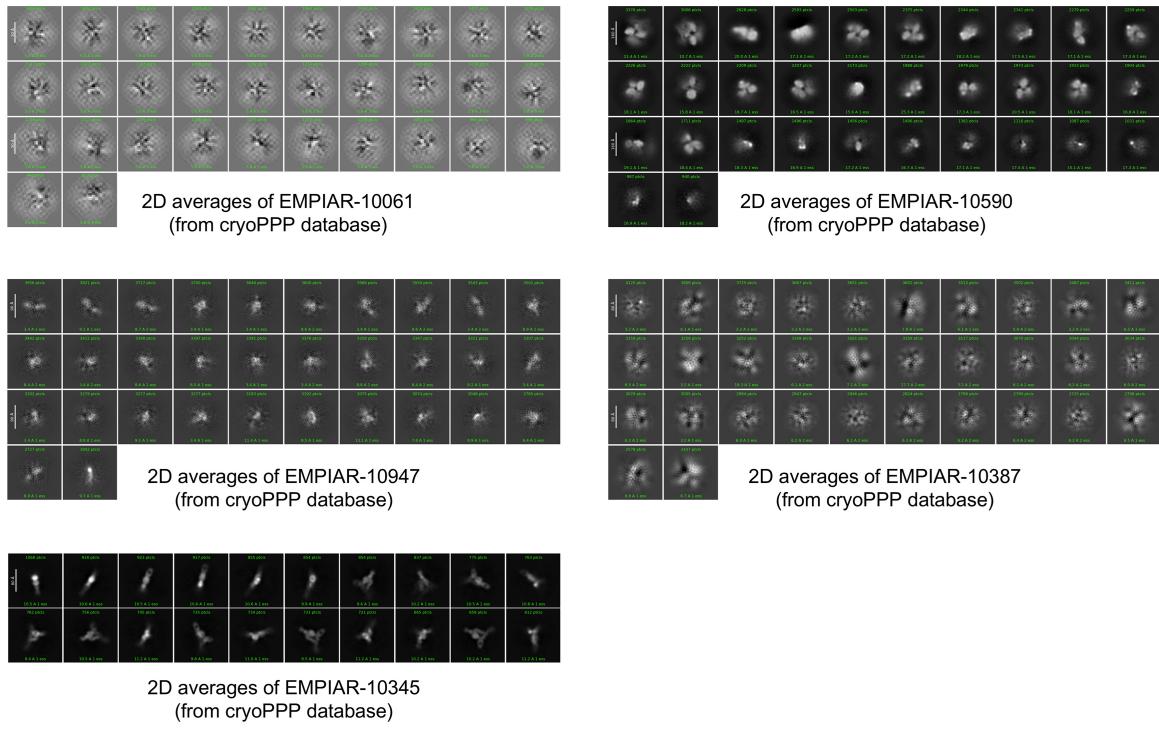
971 **models.** (a-c) The detailed architectures of the prediction head (a), the projection head (b),
 972 and the classifier head (c) in the AI models are illustrated. (d)(e) Scale test of the dataset
 973 (d) and backbone (e) sizes on the performance of the Cryo-IEF model. (f)(g) Different
 974 training parameters (f) and loss functions (g) on the fine-tuned performance of CryoRanker.
 975



976

977 **Extended Data Fig. 7 | Pipeline for preparing training datasets.** The pre-training dataset
978 contains particles from various sources (EMPIAR, CryoPPP, and In-house datasets). The
979 fine-tuning dataset is a subset of the pre-training one, where particles were processed by 2D
980 classification in CryoSPARC and manually assigned quality scores based on the clarity of
981 the class averages.

982



983

984 **Extended Data Fig. 8 | 2D class averages of datasets excluded for fine-tuning.** Five
985 datasets (EMPIAR-100061, -10590, -10947, -10387, and -10345) from CryoPPP were
986 excluded from the fine-tuning dataset due to bad qualities as visualized by 2D classification
987 results.

988

989 **Extended Data Table 1 | Summary of cryo-EM data from EMPIAR**

EMPIAR ID								
10650	11188	11265	11549	11631	11149	10381	10357	
11738	11732	11535	11530	11202	10881	10848	10683	
10469	10192	10969*	10256*	10482*	10126*	10127*	11233*	
10255*	10278*	10786*	10395*	10970*	10083*	10703*	10483*	
10396*	11120*	10254*	10280*	10722*				

990 * Provided with particle images in EMPIAR.

991

992 **Extended Data Table 2 | Summary of cryo-EM data from CryoPPP database**

EMPIAR ID							
10075	10081	10093	10059	10184	10289	10061	10017
10028	10005	10406	10291	10389	10444	11183	11057
10532	10669	10576	11051	10737	10760	10852	11056
10947*	10387*	10345*	10061*	10590*			

993 * Excluded from the fine-tuning dataset.